OULUN YLIOPISTO
UNIVERSITY of OULU

OULU BUSINESS SCHOOL

Miikka Rytty

AFFINE MULTI-FACTOR SHORT-RATE MODELS IN TERM STRUCTURE MODELING

Master's Thesis
Department of Finance
May 2019

UNIVERSITY OF OULU
Oulu Business School

ABSTRACT OF THE MASTER'S THESIS

Unit			
Department of Finance			
Author Miikka Rytty		Supervisor Professor Jukka Perttunen	
Title Affine multi-factor short-rate models in term structure modeling			
Subject Finance	Type of the degree Master's Thesis	Time of publication May 2019	Number of pages 177

Abstract

This thesis gives an overview of short-rate models in term structure modeling of interest rates. The focus is in simple preference-free models with affine term structures. The thesis also shows how these models can be extended to cover credit spreads over the risk-free interest rate. The empirical section analyzes how well these models can be applied to recent interest rate data. While short-rate models are still used, more recent market models have eclipsed them in pricing of complex derivatives. The previous literature in short-rate modeling has been mainly conducted before the financial crisis of 2007-08 and there is very little literature on comparing how well these models perform in the current market structure which features negative interest rates.

The first chapters give an overview of arbitrage-free pricing methodology of contingent claims and short-rate modeling of interest rates and credit spreads. These chapters present analytical pricing formulas for zero-coupon bonds with and without credit risk and semi-analytical pricing method for options on zero-coupon bonds in simple preference-free affine multi-factor short-rate models.

The main finding of the empirical study shows that single-factor models do not fit the recent market data. For multi-factors models, the results were not conclusive. The calibration of multi-factor models is very hard multi-dimensional optimization problem with heavy computational burden. While the quality of the multi-factor model calibrations was mostly lacking, the mixed results suggest that insufficient computing power might be cause. The rationale for this conclusion was that the calibration algorithm could not replicate previous calibration results when a different starting population was used in optimization. It seems that there were not enough computational resources to guarantee that stochastic optimization algorithm was able to find optimal parameter values.

Based on the findings of the empirical study, it seems that multi-factor short-rate models with affine term structure can be used in term structure modeling but with caveats. The whole discount curve from over-night rate to the maturity of 30 years seems to be too complex for these models but shorter sections worked much better and optimization and the computational burden may not be ignored in a more serious calibration attempt.

Keywords
Short-rate, interest rate modeling, preference-free affine models, credit risk

Additional information

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Overview of the arbitrage-free asset pricing

What is an appropriate price of an asset? Financial assets, which are contracts over other assets, are often homogeneous and standardized. Intuitively their prices should mainly depend on investors' preferences over the distributions of future returns. Investors have different preferences. Risk appetite, regulatory requirements, investment horizons and business concerns differentiates them. Secondly, history has showed that predicting the future is difficult and, as a corollary, detailed prediction of price behavior should be also hard. Without auguries, the derivation of exact probabilities is impossible.

According to Cochrane (2009), "asset pricing theory tries to understand the prices or values of claims to uncertain payments (p. xiii)". Asset pricing theory has two main approaches, absolute and relative pricing. Absolute pricing tries to model sources of economic risks and/or underlying preferences, and derive prices from these. The canonical examples are CAPM and other equilibrium models. In contrast, relative pricing does not try the model the whole investment universe. In relative prices, some asset prices are exogenous and other assets are priced relative to these. Black-Scholes–model (Black and Scholes 1973) is canonical example of this approach. The demarcation of these approaches is not clear-cut. CAPM assumes the equilibrium prices as given and Black-Scholes–model makes a fundamental assumption about the distribution of asset returns. (Cochrane 2009, pp. xiii–xiv)

The relative pricing approach is often called as arbitrage pricing methodology. The common theme in these models is that we make an assumptions about the underlying distribution of the exogenous prices. If the market is arbitrage-free, then these exogenous prices contain meaningful information about probabilities for different outcomes probability distribution. Thus we do not try to derive actual probability distribution, but we construct a new probability measure, which is commonly called as the equivalent martingale measure (EMM). These implied distributions are then used to construct replicating portfolios. By a construction of a perfect hedge, preferences do not affect the price of a replicating portfolio.

The seed of the arbitrage pricing theory is usually attributed to the thesis by Bachelier (1900). But the research by Jovanovic and Le Gall (2001) suggest that Bacheliers work was antecedent by another French Regnault (1863). The insight of these early pioneers was not try to outguess the market but the idea was model the price movements by Brownian motion[1] and use the distribution to price the derivative contract. Although the work of Bachelier was temporarily forgotten, it resurfaced in the 1950's and influenced a host of work (Samuelson 1973). Thus Regnault and Bachelier can be also seen as the forefathers of the efficient-market hypothesis which was largely developed in modern sense by Fama (1965) and Samuelson (1965a).

These early modern attempts to price options were not satisfactory even to the authors who derived them[2] until the seminal work by Black and Scholes (1973). The major insight in this and subsequent work was that, in certain idealized frictionless market model, the cash flow of the option could be perfectly replicated by trading the underlying stock and a risk-free bond. Since there was no practical difference in executing this trading strategy and owning the option, there was a unique no-arbitrage price for the option. This price depends neither the risk aversion of the inverstor nor his views of the market (expect for the volatility of stock price).

The mathematical formulation in Black and Scholes (1973) was lacking. According to Musiela and Rutkowski (2005, p. 129), it was Bergman (1982) who first noted that the trading strategy used by Black and Scholes (1973) was neither risk-free nor self-financing. The modern arbitrage asset pricing theory was formalized by Harrison and Kreps (1979) in discrete time and Harrison and Pliska (1981) in continuous time (see also Harrison and Pliska (1983)). Several authors have expanded these works and Delbaen and Schachermayer (1998) (and references within) present one version of the theory in very general a semi-martingale setting.

1.1.1. Arbitrage-free interest rate models

The demand for accurate pricing models for interest rate derivatives is great, as the positions are large. According to BIS (2018), the outstanding amount of over the counter interest rate derivatives was over 436.8 trillion USD. However, since many of the positions are overlapping, gross market value gives a better estimate of the actual size of derivative positions. This gross market value was about 6.84 trillion. Figures 1.1 and 1.2 show the development OTC derivative market segments.

[1]The problem with Brownian motion is that with non-zero probability the price will be negative given enough time. A more valid approach is to use geometric Brownian motion which do not have this problem. But geometric Brownian motion is still continous so it will not model the jumps in the price process.

[2]See, for example, the lamentation in Samuelson (1965b).

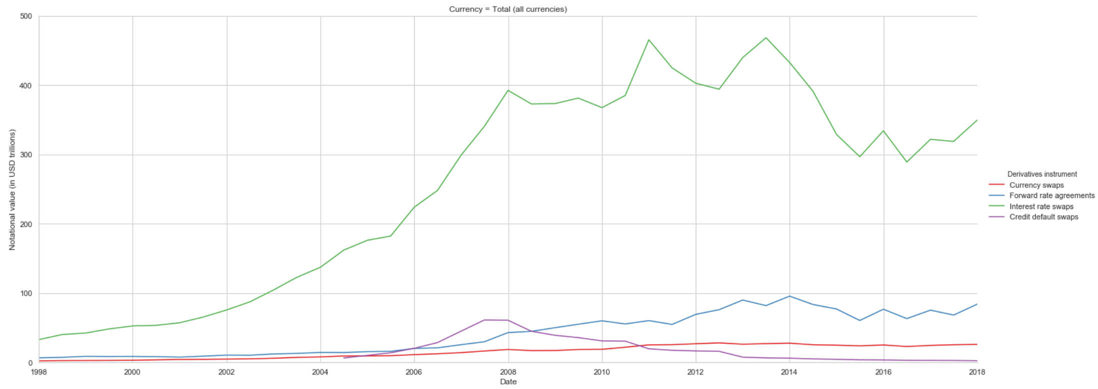


Figure 1.1: OTC derivatives notional amount outstanding. Source: BIS 2018

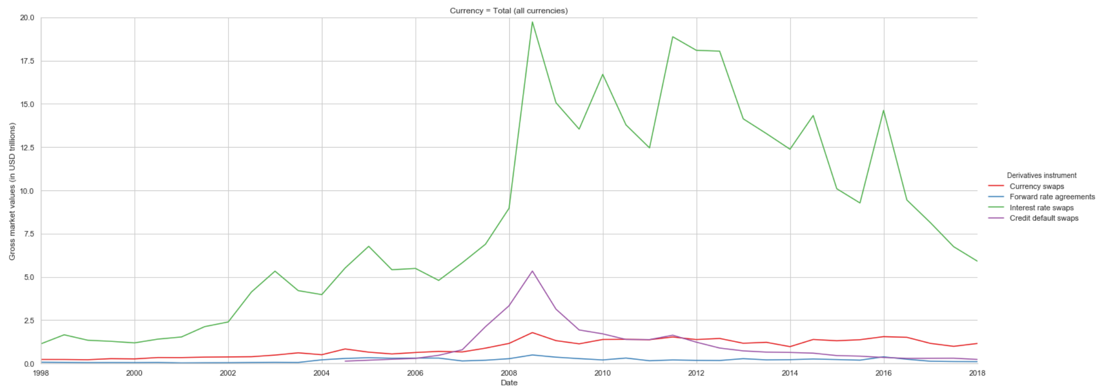


Figure 1.2: OTC derivatives gross market values. Source: BIS 2018

A vast literature exist to explain the mechanics driving the term-structure of interest rates. Short-rate modeling began under this framework and was based on macroeconomical argumentation. According to Duffie (2010, p. 161), the earliest example of markovian term-structure model is by Pye (1966). His approach influenced Merton (1974), a paper which contains an example of gaussian short-rate model. The first relevant short-rate model is the famous Vašíček–model by Vašíček (1977). Another widely used short-rate model is CIR–model by Cox, Ingersoll Jr, and Ross (1985). Since the original formulation of these models are based on economic equilibrium argumentation, they are often called either equilibrium or fundamental models and risk-preferences have to be explicitly formulated and they influences prices. Although fundamental models may have economically sound argumentation, their basic problem is practical impossibility to fit them to the observed interest-rate structure.

These early models can be also cast in arbitrage pricing framework, but we lose the economic justification behind the interest rate process. Arbitrage-free short-rate modeling just assumes that the interest rates are generated by a given exogenous stochastic process, a short-rate. This rate is mathematically modeled by an Itó-process. Observed rates are then stochastic integrals of this short rate. By this approach, pricing of interest rate derivatives leads to solving stochastic differential equations. When the short-rate model has constant parameters, then the observed interest-rate structure may not be matched. Ho and Lee (1986) introduced a short-rate model with time-varying parameter, which can be chosen to fit the observed term-structure. Vašíček–model was extended with time-varying parameters in J. Hull and White (1990).

Since short-rate models capture only a point, they have hard time capturing the complex dynamics of the term-structures. For example, we later show that in single-factor affine term-structure models rates of different maturities are perfectly correlated. An alternative to the short-rate approach is to directly model the entire term structure of interest rates. An early successor was HJM–framework (introduced in Heath, Jarrow, and Morton 1990 and Heath, Jarrow, and Morton 1992), which models the entire forward rate process. It should be noted Ho-Lee and Hull-White models can be cast as special cases of HJM model. A major problem for general HJM models is that these are not necessary markovian (Ritchken and Sankarasubramanian 1995).

Nowadays so called market models are widely used in pricing. According to Wu (2009, p. 182), traders had assumed that LIBOR and swap rates were log-normal processes and used Black's formula[3] to quote volatility when pricing caps and swaptions since the early 1990's. Theoretical justification for these practices were finally found in 1997 by a series of articles by Brace, Gatarek, and Musiela 1997, Miltersen, Sandmann, and Sondermann (1997) and Jamshidian (1997). Market models use LIBOR (or other market interest rates) as the fundamental objects and assume that they can be modeled as log-normal processes. Thus they are often called as LIBOR models. The main feature of market models is that cap and swaption pricing is given by Black's formula and they can be made to fit given term-structure and volatility structures. Since the LIBOR rates follow log-normal rates, these models have to be extended to handly negative interest rate. This is often done by modeling a shifted log-normal process. A popular extension of market model is SABR volatility model by Hagan, Kumar, Lesniewski, and Woodward (2002).

The global financial crisis of 2007-08 has had a major impact on interest-rate modeling. One of the significant additions is the multi-curve framework (Mercurio (2009)). Before the crisis, the spread between overnight indexed swap (OIS) and LIBOR curves were minimal and LIBOR curve was both the discounting and forward rate generating curve. During the crisis, this spread widened dramatically and afterwards it has been necessary the model these curves separately.

[3] See Black (1976) and section 3.2.3.

1.1.2. Intensity based modeling of credit risk

An early example of credit risk modeling is Merton (1974), which uses Black-Scholes–model to price corporate debt subject to credit risk. In Merton's model corporate assets are assumed to follow geometric brownian motion and the corporate debt consists of single zero-coupon bond. Now the equity can be seen as a call option on corporate assets with the strike price of the face value of a bond at the maturity. Thus the equity price is given by the Black-Scholes formula and the bond price is the difference of asset and equity values. Merton (1974) is extended by proprietary Moody's KMV model.

An alternative to the structural credit models are dynamic models. One approach is to use so called intensity based modeling of default. In this framework, the default time is a stopping time with a intensity process[4]. This intensity process may be influenced by properties of economy or the underlying entity. In the arbitrage-free settings intensity may be modeled by either Poisson or Cox processes. This approach was pioneered by Artzner and Delbaen (1995), Jarrow and Turnbull (1995) and Lando (1998).

The popularity intensity based modeling is due to synergies with short-rate models of interest rates. When set-up correctly, the default intensity process is the credit spread process. Then the pricing of debt and related instruments can be made using the machinery developed for the short-rate processes. J. Schönbucher (2001) has extended market models to cover credit intensity modeling.

After the global financial crisis of 2007-08, the credit risk modeling has gained importance. For example, Basel III framework requires that the prices of unsecured derivative positions has to be corrected with CVA[5], which accounts for the counterparty credit risk (BIS 2015).

1.2. Overview of the thesis

The purpose of this thesis is to have a rough overview arbitrage-free pricing methodology and affine short-rate processes used in interest rate modeling and credit risk. Although short-rate models have been eclipsed by market models, they still have their uses in risk management, portfolio management and scenario planning.

As the short-rate models were developed before global financial crisis of 2007-08, we test how well they can be fitted to the post-crisis interest rate data. We also try to test calibrate a combined short-rate and credit spread model to post-crisis bond price data.

[4] See chapter 5 and section A.5.

[5] CVA is credit value adjustment and pricing has also account DVA, debit value adjustment.

Chapter 2

Chapter 2 gives a basic overview of the common interest rates and financial instruments.

Chapter 3

Chapter 3 gives a very brief introduction to arbitrage pricing theory. We first develop discrete one period model in order to highlight the basic concepts of arbitrage pricing such as justification behind martingale measures and laws of asset pricing. This treatment is based on Björk (2004, pp. 5–34). All the proofs are detailed as we feel that they give insight to martingale measures.

After that the arbitrage pricing in continuous markets is overviewed. This treatment is not rigorous. Measure theoretical justifications and arguments are simply omitted although some basic results are presented in Appendix A.

While not strictly necessary for the empirical work in this thesis, the arbitrage pricing theory is essential for understanding the peculiarities in interest rate and credit spread modeling.

Chapter 4

Chapter 4 starts with the derivation of term-structure equation and the overview of the fundamental and preference-free models. We have a brief overview of the single-factor Vašíček and CIR–models.

We also cover multi-factor affine $A(M,N)$–models, which are a combination of M Vašíček and $N-M$ CIR–models. The Gaussian processes can be chosen to be correlated but square-root processes has to be uncorrelated. The theory of affine term-structure models was developed for example in Brown and Schaefer (1994), Duffie and Kan (1994) and Duffie and Kan (1996). The main advantage of these affine models is that they have analytical bond prices which make calibration by term-structure easy. By dynamic extension shifting[6], we even could make sure that term-structures fit perfectly. Many of these models also have analytical bond option prices, which allows us to price caps and floors and for single factor models, this allows also easy pricing of swaptions. Therefore they can be easily fit to volatility structures also. Even if the model does not have a bond option pricing formula, we may use Fourier transformations to have semi-analytical pricing of bond options[7].

[6]See Brigo and Mercurio (2001) and section 4.4.
[7]See Heston (1993) and section 4.5.

All the simple preference-free models models[8] covered in Chapter 4 see implementation in the empirical work in Chapter 6.

Chapter 5

Chapter 5 is an overview of default intensities in credit risk modeling. We show how a Cox process[9] can be used to extended the machinery of short-rate interest rate modeling to model credit spreads. In order to do this, we show how to price discounted cash flows that depend on timing of a default event, where both short-rate and default intensity are stochastic in nature. We develop model-independent pricing formulas of defaultable zero-coupon bonds. Although credit default swaps are not covered in the emperical work of Chapter 6, we also show how easily these methods can be used to the price credit default swaps.

There is also an exposition covering on how the multi-factor affine $A(M,N)$–models of Chapter 4 can be extended to model both interest rates and default intensities. In the empirical chapter, we try to calibrate these models to market data.

Chapter 6

Chapter 6 presents the methods and results from the empirical work done in this thesis. We test some of the methods develop in earlier chapters. In order to find suitable initial starting value for descending optimization algorithm, we tried to scan the problem space with a differential evolution (DE) algorithm. Due to computational limitations, the sample sizes were small compared to the dimension of the problem space. As such, the algorithm did not produce consistent results.

Simple affine models were fitted to various interest rate and yield curves with and without default risk. Fittings of these curves proved to be challenging for the models as the maturities ranged from overnight rates (or 6-month rates) to 30-year rates. No model provided satisfactory fitting in every case. Fitting done by models with credit risk component proved bad overall. However, due to the problems with optimization algorithm, we can not decisively rule that the models will provide bad fits with the used data.

Dynamic fitting of Euribor rates ranging from 1-week rate to 1-year rate was also attempted to single factor models. Depending on the time-period, the fits ranges from horrible to rather satisfactory. As the short end of the rate curve has been rather flat after 2014, calibrated curves had very little errors.

[8] See section 4.1 for introduction to these models.

[9] Cox process is also called as doubly stochastic process, since in addition to the stochastic stopping time, the time dependent intensity process is also stochastic.

Due to data and computational limitations, fitting to the volatility structures was not attempted.

The code with the Jupyter notebook used in analysis can be found at `https://github.com/mrytty/gradu-public` (Rytty (2019)).

Appendices A and B

Appendix A is a brief review of mathematical machinery and methods needed in the earlier chapters.

Appendix B presents some additional data charts and tables from the empirical work.

1.3. Fixed notation

The following notation will be fixed.

$\Delta(t,T)$ Day count convention fraction between the dates $t < T$.

$r(t)$ Short-rate at the time t.

$B(t)$ The value of an idealized bank account at the time t.

$D(t,T)$ Stochastic discount factor between the dates t,T.

$p(t,T)$ The price of T-bond (zero coupon bond with the maturity T) at the time t.

$L(t,T)$ Simple spot rate between dates t and T

$L^k(t,T)$ k-times compounded simple spot rates between dates t and T

$\mathrm{FRA}(t,T,S,K)$ The price of a forward rate agreement at the time t, where K is the fixed price and the interest is paid between the dates $T < S$.

$L(t,T,S)$ Simple forward rate at the time t between dates T and S

$f(t,T)$ Instantaneous forward rate at the time t for date T.

$\mathrm{Fut}(t,T,S)$ Futures rate at the time t between dates T and S.

$d(t,T)$ The price of defaultable T-bond (zero coupon bond with the maturity T) at the time t.

$\mathrm{ZBC}(t,S,T,K)$ The price of call option at the time t maturing at S written on T-bond.

ζ The time of a default.

$\mathrm{L_{GD}}$ The loss given the default (LGD) of a contract.

$\mathrm{R_{EC}}$ The recovery value of a contract.

$\mathrm{CDS_{pro}}(t)$ The value of a protection leg of a credit default swap.

$\mathrm{CDS_{pre}}(t,C)$ The value of a premium leg of a credit default swap with a coupon rate C.

2. IDEALIZED RATES AND INSTRUMENTS

In this chapter we review idealized versions of interest rates and derivatives of interest rates. By idealized, we mean that the instruments are simplied for analytical purposes. For example, there is no lag between trade and spot date or expiry and delivery date. Neither we do not use funding rate that is separate form the market rates. The treatment is standard and is based mainly on Brigo and Mercurio (2007, pp. 1–22) unless otherwise noted.

In the following, we assume that $0 < t < T$ are points of time and $\Delta(t,T) \in [0,\infty)$ is the day count convention between the points t and T. We explicitly assume that $\Delta(t,T) \approx T - t$ when $t \approx T$.

2.1. Fundamental rates and instruments

2.1.1. Short-rate, idealized bank account and stochastic discount factor

When making calculations with idealized bank account, it is customary to assume that the day count-convention $\Delta(t,T) = T - t$ as this will simplify the notation. An idealized bank account is an instrument with the value

$$B(t) = \exp\left(\int_0^t r(s)\mathrm{d}s\right) \tag{2.1.1}$$

where $r(t)$ is the short-rate rate. The short-rate $r(t)$ may be non-deterministic but we assume that it is smooth enough so that the integral can be defined in some useful sense. We note that $B(0) = 1$. If $\delta > 0$ is very small and $r(t)$ is a smooth function, then

$$\int_t^{t+\delta} r(s)\mathrm{d}s \approx r(t)\delta \tag{2.1.2}$$

and we see that the first-order expansion of exponential function yields

$$B(t+\delta) \approx B(t)(1 + r(t)\delta). \tag{2.1.3}$$

Thus the short-rate can be seen as continuous interest rate intensity. Short-rate is purely theoretical construction which can be used to price financial instruments.

Now we may define a stochastic discount factor $D(t,T)$ from time t to T as

$$D(t,T) = \frac{B(t)}{B(T)} = \exp\left(-\int_t^T r(s)\mathrm{d}s\right). \tag{2.1.4}$$

If $r(t)$ is a random variable, then $B(t)$ and $D(t,T)$ are stochastic too.

2.1.2. Zero-coupon bond

A promise to pay one unit of currency at time T is called a T-bond. We shall assume that there is no credit risk for these bonds. We further assume that the market is liquid and bond may be freely bought and sold at the same price, furthermore short selling is allowed without limits or extra fees. The price of this bond at time t is denoted by $p(t,T)$ and so $p(t,T) > 0$ and $p(T,T) = 1$.

As $D(0,t)$ is guaranteed to pay one unit of currency at the time t, we see that in this case $D(0,t) = p(0,t)$. We note that if the short-rate $r(t)$ is deterministic, then

$$D(0,t) = \frac{1}{B(t)} \tag{2.1.5}$$

is deterministic too. We see that if short-rate $r(t)$ is deterministic, then $D(t,T) = p(t,T)$ for all $0 \leq t \leq T$. But this does not hold if $r(t)$ is truly stochastic.

2.1.3. Simple spot $L(t,T)$ and k-times compounded simple spot rate

The simple spot rate $L(t,T)$ is defined by

$$L(t,T) = \frac{1 - p(t,T)}{\Delta(t,T)p(t,T)}, \tag{2.1.6}$$

which is equivalent to

$$1 + \Delta(t,T)L(t,T) = \frac{1}{p(t,T)}. \tag{2.1.7}$$

For $k \geq 1$, the k-times compounded interest rate from t to T is

$$L^k(t,T) = \frac{k}{p(t,T)^{\frac{1}{k\Delta(t,T)}}} - k, \tag{2.1.8}$$

which is equivalent to

$$p(t,T)\left(1+\frac{L^k(t,T)}{k}\right)^{k\Delta(t,T)} = 1. \tag{2.1.9}$$

As

$$(1+\frac{x}{k})^k \longrightarrow \mathrm{e}^x \tag{2.1.10}$$

when $k \longrightarrow \infty$, then

$$\left(1+\frac{L^k(t,T)}{k}\right)^{k\Delta(t,T)} \longrightarrow \mathrm{e}^{\Delta(t,T)r(t,T)}, \tag{2.1.11}$$

where $L^k(t,T) \longrightarrow r(t,T)$ when $k \longrightarrow \infty$.

2.1.4. Forward rate agreement

A forward rate agreement (FRA) is a contract that pays

$$\Delta_K(t,T)K - \Delta(t,T)L(t,T) \tag{2.1.12}$$

at the time T. Here we assume that the contact is made at the present time 0 and $0 < t < T$, but this assumption is made just to keep the notation simplier. Here K is an interest rate that is fixed at time 0, Δ_K is the day count convention for the this fixed rate and $L(t,T)$ is the spot rate from time t to T (which might not be know at the present). The price of a FRA at the time $s \le t$ is denoted by $\mathrm{FRA}(s,t,T,K)$. Now

$$\mathrm{FRA}(t,t,T,K) = p(t,T)\left(\Delta_K(t,T)K - \Delta(t,T)L(t,T)\right). \tag{2.1.13}$$

In order to price a FRA at different times, we consider a portfolio of one long T-bond and x short t-bonds. The value of this portfolio at the present is $V(0) = p(0,T) - xp(0,t)$ and we note that the portfolio has zero value if

$$x = \frac{p(0,T)}{p(0,t)}. \tag{2.1.14}$$

At the time t, the portfolio has value

$$V(t) = p(t,T) - x \tag{2.1.15}$$
$$= p(t,T)\left(1 - \frac{x}{p(t,T)}\right) \tag{2.1.16}$$

where $p(t,T)$ is known and

$$1 + \Delta(t,T)L(t,T) = \frac{1}{p(t,T)} = y(t,T). \tag{2.1.17}$$

We define $K^*(x) = x^{-1}$. Thus

$$1 - \frac{x}{p(t,T)} = x\left(\frac{1}{x} - \frac{1}{p(t,T)}\right) \tag{2.1.18}$$
$$= x\left(K^*(x) - y(t,T)\right) \tag{2.1.19}$$

and this implies that

$$V(t) = xp(t,T)\left(K^*(x) - y(t,T)\right). \tag{2.1.20}$$

Without arbitrage

$$V(T) = x\left(K^*(x) - y(t,T)\right) \tag{2.1.21}$$
$$= x\left(K^*(x) - 1 - \Delta(t,T)L(t,T)\right) \tag{2.1.22}$$

We note that at the time 0, $K^*(x)$ is a known yield but $y(t,T)$ is an unknown yield if $p(t,T)$ is not deterministic. Now if

$$K = \frac{1}{\Delta_K(t,T)}\left(K^*(x) - 1\right) \tag{2.1.23}$$
$$= \frac{1}{\Delta_K(t,T)}\left(\frac{1}{x} - 1\right) \tag{2.1.24}$$

the given portfolio can be used to replicate the cash flows of the FRA and

$$x\mathrm{FRA}(s,t,T,K) = V(s). \tag{2.1.25}$$

If

$$x = \frac{p(0,T)}{p(0,t)} \tag{2.1.26}$$

then $V(0) = 0$ and

$$K = \frac{1}{\Delta_K(t,T)}\left(\frac{p(0,t)}{p(0,T)} - 1\right) \tag{2.1.27}$$

$$= \frac{\Delta(t,T)}{\Delta_K(t,T)} L(t,T). \tag{2.1.28}$$

We see that the forward rate and the rate that defines FRA with zero present value are essentially the same. Thus we define that the forward rate at the time t from time T to S is

$$L(t,T,S) = \frac{1}{\Delta(T,S)}\left(\frac{p(t,T)}{p(t,S)} - 1\right). \tag{2.1.29}$$

Since $\Delta(T,S) \approx S - T$ when $T \approx S$, we have that

$$L(t,T,S) = \frac{1}{\Delta(T,S)}\left(\frac{p(t,T)}{p(t,S)} - 1\right) \tag{2.1.30}$$

$$\approx \frac{1}{p(t,T)} \frac{p(t,T) - p(T,S)}{S - T} \tag{2.1.31}$$

and therefore

$$L(t,T,S) \longrightarrow -\frac{1}{p(t,T)} \frac{\partial p(t,T)}{\partial t} \tag{2.1.32}$$

$$= -\frac{\partial \log p(t,T)}{\partial T} \tag{2.1.33}$$

when $S \to T^+$ under the assumption that the zero curve $p(t,T)$ is differentiable. We now define that the instantaneous forward rate at the time t is

$$f(t,T) = -\frac{\partial \log p(t,T)}{\partial T}. \tag{2.1.34}$$

Now since $p(t,t) = 1$,

$$-\int_t^T f(t,s)\mathrm{d}s = \int_t^T \partial \log p(t,s)\mathrm{d}s \tag{2.1.35}$$

$$= \log p(t,T) - \log p(t,t) \tag{2.1.36}$$

$$= \log p(t,T) \tag{2.1.37}$$

meaning that

$$p(t,T) = \exp\left(-\int_t^T f(t,s)\mathrm{d}s\right). \tag{2.1.38}$$

2.2. Interest rate instruments

2.2.1. Fixed leg and floating leg

A leg with tenor $t_0 < t_1 < t_2 < \ldots < t_n = T$ and coupons $c_1, c_2, \ldots, c_n$ is an instruments that pays c_i at the time t_i for all $1 \leq i \leq n$. The coupons may be functions of some variables. Thus a is a portfolio of n zero-coupon bonds with maturities coinciding with tenor. It has has present value of

$$\sum_{i=1}^{n} c_i p(t,t_i) 1_{\{t \geq t_i\}} \tag{2.2.1}$$

at the time t.

A floating leg with a unit principal has coupons defined by $c_i = \Delta_1(t_{i-1},t_i)L(t_{i-1},t_i)$, where L is a reference rate for a floating. It has a present value of

$$\begin{aligned}
PV_{\text{float}}(t) &= \sum_{i=1}^{n} p(t,t_i)\Delta_1(t_{i-1},t_i)L(t_{i-1},t_i) && (2.2.2)\\
&= \sum_{i=1}^{n} p(t,t_i)\left(\frac{1}{p(t_{i-1},t_i)} - 1\right) && (2.2.3)\\
&= \sum_{i=1}^{n} p(t,t_{i-1})p(t_{i-1},t_i)\left(\frac{1}{p(t_{i-1},t_i)} - 1\right) && (2.2.4)\\
&= \sum_{i=1}^{n} \left(p(t,t_{i-1}) - p(t,t_{i-1})p(t_{i-1},t_i)\right) && (2.2.5)\\
&= \sum_{i=1}^{n} \left(p(t,t_{i-1}) - p(t,t_i)\right) && (2.2.6)\\
&= p(t,t_0) - p(t,t_n) && (2.2.7)
\end{aligned}$$

and especially $PV_{\text{float}}(t_0) = 1 - p(t,t_n)$.

If the coupons are $c_i = K\Delta_0(t_{i-1},t_i)$ for a fixed rate K, then we call it as a fixed leg with a unit principal. It has a present value

$$PV_{\text{fixed}}(t) = K\sum_{i=1}^{n} \Delta_0(t_{i-1},t_i)p(t,t_i). \tag{2.2.8}$$

2.2.2. Coupon bearing bond

A coupon bearing bond with floating coupons and a unit principal is combination of a floating leg and payment of one currency unit coinciding with the last tenor date. Thus it has present value of

$$PV_{\text{floating bond}}(t) = p(t,t_0) \tag{2.2.9}$$

and especially $PV_{\text{floating bond}}(t_0) = 1$.

Similarly a coupon bearing bond with fixed coupons and a unit principal is combination of a fixed leg and payment of one currency unit coinciding with the last tenor date. It has a present value of

$$\begin{align} PV_{\text{fixed bond}}(t) &= p(t,t_n) + K\sum_{i=1}^{n} \Delta_0(t_{i-1},t_i)p(t,t_i) \tag{2.2.10} \\ &= p(t,t_n) + PV_{\text{fixed}}(t). \tag{2.2.11} \end{align}$$

2.2.3. Vanilla interest rate swap

A vanilla payer interest rate swap (IRS) is a contract defined by paying a fixed leg and receiving a floating leg. A vanilla receiver interest rate swap (IRS) is a contract defined by paying a floating leg and receiving a fixed leg. The legs may have different amount of coupons. Also the coupons dates and day count conventions may not coincide. If K is the common rate for the fixed leg and both legs have the same notional value, then a payer IRS has the present value of

$$\sum_{i=1}^{m} p(t,t_i')\Delta_1(t_{i-1}',t_i')L(t_{i-1}',t_i') - K\sum_{i=1}^{n} p(t,t_i)\Delta_0(t_{i-1},t_i) \tag{2.2.12}$$

where $t_0' < t_1' < t_2' < \ldots < t_m'$ are the coupon times for the floating leg. A par swap is a swap with present value of zero and the fixed rate for a par swap is

$$K = \frac{\sum_{i=1}^{m} p(t,t_i')\Delta_1(t_{i-1}',t_i')L(t_{i-1}',t_i')}{\sum_{i=1}^{n} p(t,t_i)\Delta_0(t_{i-1},t_i)} \tag{2.2.13}$$

It is easy to see that if the both legs have same underlying notional principal and coupon dates are the same, then the swap is just a collection of forward rate agreements with a fixed strike price. A vanilla payer IRS let the payer to hedge interest rate risk by converting a liability with floating rate payments into fixed payments.

2.2.4. Overnight indexed swap

At the end of a banking day, banks and other financial institutions may face surplus or shortage of funds. They may lend the excess or borrow the shortfall on overnight market. Overnight lending rate is often regarded as a proxy for risk-free rate. In Euro area, European Central Bank calculates Eonia, which is a weighted average of all overnight unsecured lending transactions in the interbank market.

Overnight indexed swap (OIS) is a swap where a compounded reference overnight lending rate is exchanged to a fixed rate.

2.2.5. Call and put option and call-put parity

A European call (put) option gives the buyer the right but not an obligation to buy (sell) a designated underlying instrument from the option seller with a fixed price at expiry date. Thus a call option on T-bond with strike price K and maturity $S < T$ has the final value

$$\mathrm{ZBC}(S,S,T,K) = (p(S,T) - K)^{+} \tag{2.2.14}$$

and the corresponding put option has the final value

$$\mathrm{ZBP}(S,S,T,K) = (K - p(S,T))^{+}. \tag{2.2.15}$$

A portfolio of long one call and short one put option on a same T-bond with identical strike price K and maturity S has final value of

$$(p(S,T) - K)^{+} - (K - p(S,T))^{+} = p(S,T) - K. \tag{2.2.16}$$

Therefore, without any arbitrage, we have the so called call-put–parity

$$\mathrm{ZBC}(t,S,T,K) - \mathrm{ZBP}(t,S,T,K) = p(t,T) - p(t,S)K \tag{2.2.17}$$

holds for all $t \le S$.

2.2.6. Caplet, cap, floorlet and floor

In order to keep notation simplier, we assume that the present is 0 and $0 < t < T$. A caplet is an interest rate derivative in which the buyer receives

$$(L(t,T) - K)^{+} \tag{2.2.18}$$

at the time T, where $L(t,T)$ is some reference rate and K is the fixed strike price. The fixing is done at when the contract is made.

Suppose that a firm must pay a floating rate L. By buying a cap with strike K against L, the firm is paying

$$L-(L-K)^{+}=\min(L,K) \tag{2.2.19}$$

meaning that the highest rate will pay will be the strike rate K. Thus caps may be used to hedge interest rate risk.

Now

$$L(t,T)-K=\frac{1}{\Delta(t,T)}\left(1+\Delta(t,T)L(t,T)-K^{*}\right) \tag{2.2.20}$$

$$=\frac{1}{\Delta(t,T)}\left(\frac{1}{p(t,T)}-K^{*}\right) \tag{2.2.21}$$

where $K^{*}=1+\Delta(t,T)K$. Thus the value of a caplet at the time t is

$$p(t,T)\left(L(t,T)-K\right)^{+}=\frac{p(t,T)}{\Delta(t,T)}\left(\frac{1}{p(t,T)}-K^{*}\right)^{+} \tag{2.2.22}$$

$$=\frac{1}{\Delta(t,T)}\left(1-p(t,T)K^{*}\right)^{+} \tag{2.2.23}$$

$$=\frac{K^{*}}{\Delta(t,T)}\left(\frac{1}{K^{*}}-p(t,T)\right)^{+}. \tag{2.2.24}$$

But this is the price of $\frac{K^{*}}{\Delta(t,T)}$ put options on a T-bond with strike price $\frac{1}{K^{*}}$ at the time of strike t. Thus we can price a caplet as a put option on a bond. As the price of a cap contains optionality, we must model the interest rates in order to price it.

A cap is a linear collection of caplets with the same strike price.

A floorlet is an derivate with the payment

$$\Delta(t,T)\left(K-L(t,T)\right)^{+} \tag{2.2.25}$$

at the time T, where $L(t,T)$ is some reference rate with day-count convention $\Delta(t,T)$ and K is the fixed strike price. Similarly a floor is a linear collection of floorlets with the same strike price. We can price a floorlet is the price of $\frac{K^{*}}{\Delta(t,T)}$ call options on a T-bond with strike price $\frac{1}{K^{*}}$ at the time of strike t.

2.2.7. Swaption

A swaption is an interest rate derivative that allows the owner the right but not an obligation to enter into an IRS. A payer swaption gives the owner the right to enter a payer swap (a swap paying a fixed rate while receiving floating rate). A receiver swaption gives the owner the option to initiate a receiver swaption (a swap paying a floating rate while receiving a fixed rate).

A European payer swaption is equivalent to a European put option on a coupon bearing bond. The underlying swap have the value of

$$\mathrm{Swap}(S) = PV_{\mathrm{float}}(S) - PV_{\mathrm{fixed}}(S). \tag{2.2.26}$$

at the time of the strike S. Thus

$$\begin{aligned} \mathrm{Swaption}(S) &= (PV_{\mathrm{float}}(S) - PV_{\mathrm{fixed}}(S))^{+} && (2.2.27)\\ &= (1 - p(t,t_n) - PV_{\mathrm{fixed}}(S))^{+} && (2.2.28)\\ &= (1 - PV_{\mathrm{fixed\ bond}}(S))^{+}. && (2.2.29) \end{aligned}$$

We see that a swaption is a european put option on a fixed rate coupon bond. The coupon rate is the fixed rate of the underlying swap and strike price is the principal of the bond and the underlying swap.

In some cases we may price a swaption as a portfolio of options on zero-coupon bond. This trick was introduced in Jamshidian 1989. We now denote the price of a zero coupon bond as a function of a short rate $p(t,T,r)$. We consider a put option with maturity S and strike price K on a bond with coupond c_i occuring at times t_i, $i = 1,2,\ldots,n$. Let r^* be the rate with the property

$$K = \sum_{i=1}^{n} c_i p(S,t_i,r^*). \tag{2.2.30}$$

Now the put option has a value

$$\left(K - \sum_{i=1}^{n} c_i p(S,t_i)\right)^{+} = \left(\sum_{i=1}^{n} c_i \left(p(S,t_i,r^*) - p(S,t_i,r(S))\right)\right)^{+}. \tag{2.2.31}$$

If we assume that the bond prices are uniformly decreasing function on the initial short rate, then the options will be exercised if and only if $r^* < r(S)$ and now

$$p(S,t_i,r^*) > p(S,t_i,r(S)). \tag{2.2.32}$$

for all i Otherwise all $p(S,t_i,r^*) \leq p(S,t_i,r(S))$ for all i. Thus the put option has value

$$\sum_{i=1}^{n} c_i \left(p(S,t_i,r^*) - p(S,t_i,r(S))\right)^+ \tag{2.2.33}$$

which is a portfolio of put options with maturities S on a zero coupon bonds with strike prices of $p(S,t_i,r^*)$. The assumption behind this trick assumes in essence that the prices of the zero coupon bonds moves in unison. This is satisfied by one-factor models but the assumption does not hold for multi-factor models.

Similarly, a European receiver swaption is equivalent to a European call option on a coupon bearing bond. Under the same assumption, we may disassemble a receiver swaption as a portfolio of call options on zero coupon bonds.

2.3. Defaultable instruments and credit default swaps

2.3.1. Defaultable T-bond

A defaultable T-bond with no recovery (NR) is an instrument that pays

$$d(T,T) = \begin{cases} 1, & T < \zeta \\ 0, & T \geq \zeta \end{cases} \tag{2.3.1}$$

at the time T, where ζ is the time of a default of the underlying. The price of a defaultable T-bond at the time $t < T$ is denoted by $d(t,T)$.

A defaultable T-bond with recovery of treasury (RT) has the same final payout is a defaultable T-bond with no recovery but in addition it pays $\delta p(\zeta,T)$ if $\zeta \leq T$, where $0 < \delta < 1$. Thus it was a terminal value of

$$d(T,T) = 1_{\{\zeta > T\}} + \delta 1_{\{\zeta \leq T\}}. \tag{2.3.2}$$

A defaultable T-bond with recovery of face value (RFV) has the same final payout is a defaultable T-bond with no recovery but in addition it pays δ at the default if $\zeta \leq T$, where $0 < \delta < 1$. Thus it was a terminal value of

$$d(T,T) = 1_{\{\zeta > T\}} + \delta 1_{\{\zeta \leq T\}} p(\zeta,T). \tag{2.3.3}$$

A defaultable T-bond with recovery of market value (RMV) has the same final payout is a defaultable T-bond with no recovery but in addition it pays $\delta d(\zeta,T)$ at the default if $\zeta \leq T$, where $0 < \delta < 1$. Thus it was a terminal value of

$$d(T,T) = 1_{\{\zeta > T\}} + \delta 1_{\{\zeta \leq T\}} p(\zeta, T). \tag{2.3.4}$$

2.3.2. Credit default swap

A credit default swap (CDS) is an instrument where the seller of the contract will compensate the buyer if the reference instrument or entity has a credit event such as a default. In exchange, the buyer will make periodic payments to the seller until the end of the contract or the default event. The buyer of CDS will be hedged against the credit risk of the reference entity. Originally physical settlement was used. If the credit event occurs before the maturity of the CDS, then the seller is obligated to buy the underlying reference debt for face value. Since the notional value of credit default swaps may be greater than the underlying debt, physical settlement is a cumbersome process and cash settlements are held instead. In order to determine the value of a contract after the default, a credit event auction is held to determine the recovery value $\mathrm{R_{EC}}$ (ISDA 2009, BIS 2010).

Suppose that the CDS will offer protection from S to T and ζ is the time of the credit event. The protection seller has agreed to pay the buyer $\mathrm{L_{GD}} = 1 - \mathrm{R_{EC}}$ at the time ζ if $S \leq \zeta \leq T$. The protection leg of CDS has a value of

$$\mathrm{CDS_{pro}}(t) = p(t, \zeta) \mathrm{L_{GD}} 1_{\{S \leq \zeta \leq T\}} \tag{2.3.5}$$

at the time t. Let $S = t_0 < t_1 < t_2 < \ldots < t_n = T$. The premium leg will pay a coupon rate C at the times $t_1 < t_2 < \ldots < t_n$ if the credit event has not occurred. If the credit event happes, then the buyer will pay the accrued premium rate at the time of the default. The premium leg has a value of

$$\mathrm{CDS_{pre}}(t, C) = \sum_{i=1}^{n} p(t, t_i) \Delta(t_{i-1}, t_i) C 1_{\{\zeta > t_i\}} + p(t, \zeta) \Delta(t, \zeta) C 1_{\{t_s \leq \zeta \leq t_{s+1}\}} \tag{2.3.6}$$

where t_s is the last date from $t_0 < t_1 < \ldots < t_n$ before the credit event (if it occurs).

Standardized CDS contracts have quarterly coupon payments and rates are usually set to be either 25, 100, 500 or 1000 basis points. So when traded the buyer will pay

$$\mathrm{CDS_{pre}}(0, C) - \mathrm{CDS_{pro}}(0). \tag{2.3.7}$$

Earlier the coupon rate C was set so that $\mathrm{CDS_{pre}}(0, C) = \mathrm{CDS_{pro}}(0)$ and no money was exchanged at the trade.

3. AN INTRODUCTION TO ARBITRAGE PRICING THEORY

We shall now review the basic setting and results of arbitrage pricing theory.

3.1. Discrete one period model

In order to gain an insight into the basic concepts in mathematical finance, an overview of discrete one period model is given in this chapter. This treatment is based on Björk (2004, pp. 5–34) and Duffie (2010, pp. 3–12). While the context follows these sources closely, the presentation should be unique. For example, we do not assume the ecistance of a risk-free rate.. A more complete overview of discrete model with multiple time periods can be found in Musiela et al. (2005, pp. 33–85).

Let $(\Omega, \mathbb{P})$ be a discrete probability space with $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ and we assume that $\mathbb{P}(\omega) > 0$ for all $\omega \in \Omega$. The discrete one period market model is the sample space Ω coupled with the price process S. We assume that at the time $t = 0$ the price vector $S_0 \in \mathbb{R}^n$ is a known constant and at the time $t = 1$ it is a random vector $S_1 : \Omega \to \mathbb{R}^n$. Thus our model consists of one time step with n stochastic assets.

A portfolio w is a vector in $\mathbb{R}^n$. Since we do not put any restrictions on w, we assume that there are no restriction on buying or short selling of assets in this market. We denote the starting value of portfolio w by $V_0^w = w^\top S_0$ and the final value of it is a random variable $V_1^w = w^\top S_1$.

We say that the portfolio w is an arbitrage portfolio if either

1. $V_0^w < 0$ and $V_1^w \geq 0$ or
2. $V_0^w \leq 0$ and $V_1^w > 0$

$\mathbb{P}$-surely. We say that the market with starting prices is arbitrage free if there are no arbitrage portfolios. We denote

$$M = \begin{bmatrix} S_1(\omega_1) & S_1(\omega_2) & \ldots & S_1(\omega_m) \end{bmatrix}_{n \times m} = \begin{bmatrix} M_1^\top \\ M_2^\top \\ \vdots \\ M_n^\top \end{bmatrix}_{n \times m} \tag{3.1.1}$$

where $M_i^\top \in \mathbb{R}_m$ is the vector for terminal prices of ith asset. Now the absence of arbitrage is equivalent to that neither

1. $w^\top S_0 = V_0^w < 0$ and $w^\top M \geq 0$ nor
2. $w^\top S_0 = V_0^w \leq 0$ and $w^\top M > 0$

does not hold for any portfolio w. We recall that for vectors, we use $x > 0$ to denote that every component of x is positive. Likewise $x \geq 0$ denotes component-wise non-negativity.

3.1.1. The first fundamental theorem of asset pricing in discrete one period model

A state-price vector is a vector $x \in \mathbb{R}^m$ satisfying $S_0 = Mx$ and $x > 0$. This definition can be understood in the following context. If $x \in \mathbb{R}^m$ is a state-price vector and $x = \beta y$, where $|y| = 1$ and $\beta = |x|$, then $S_0 = Mx = \beta My$. where My is the expected value of S_1 under the probability measure where the state probabilities are given by the vector y. The probability vector of the original probability space $\mathbb{P}$ does not have to be a state-price vector.

It may happen that the state-price vector does not exists, but if it does, then $V_0^w = w^\top S_0 = w^\top Mx$ for all portfolios w and this implies that arbitrage does not exist. Suppose that $w \in \mathbb{R}^n$ and x is a state-price vector. Now $x > 0$ and $w^\top S_0 = w^\top Mx < 0$ implies that $w^\top M$ has a negative component. Likewise $x > 0$ and $w^\top S_0 = w^\top Mx = 0$ implies that not all elements of $w^\top M$ are positive.

We shall now prove the reverse implication using a variant of hyperplane separation theorem, which can be found in many standard textbooks for convex optimization.

Theorem 3.1.1 (Hyperplane separation theorem)**.** *Let A and B be closed convex subsets of R^s. If either of them is compact, then there exists $0 \neq x \in R^s$ such that*

$$a^\top x < b^\top x \tag{3.1.2}$$

for all $a \in A$ and $b \in B$.

Lemma 3.1.2. *The market (Ω, S) is arbitrage free if and only if a state-price vector exists.*

Proof. The argument presented here is essentially the same as the one found in Duffie (2010, p. 4). We denote

$$A = \{\ (-w^\top S_0, w^\top M) \in \mathbb{R}^{m+1} \mid w \in \mathbb{R}^n\ \} \tag{3.1.3}$$

and $C = \mathbb{R}_+^{m+1}$. Now the absence of arbitrage is equivalent to $A \cap C = \{0\}$. We need to only prove that the absence of arbitrage implies that the state-price vector does exist.

It is clear that A is a closed and convex linear subspace of $\mathbb{R}^{m+1}$. We consider the function f defined by $z \mapsto z/|z|$ for all $0 \neq z \in \mathbb{R}^{m+1}$. Since C is a convex subset, it is easy to to see that the convex closure of $B = f(C \setminus \{0\})$ is a convex and compact subset of C.

By the hyperplane separation theorem, there exists $0 \neq y \in R^{m+1}$ such that $a^\top y < b^\top y$ for all $a \in A$ and $b \in B$. Since $0 \in A$, $0 < b^\top y$ for all $b \in B$, which implies that $0 < c^\top y$ for all $0 \neq c \in C$. Coordinate vectors are in C and this means that $y > 0$. Since A is a subspace, $a \in A$ implies that $-a \in A$ and this means that $0 \leq a^\top y < \left|c^\top y\right|$ for all $a \in A$ and $0 \neq c \in C$. Thus $a^\top y = 0$ for all $a \in A$.

If $y = (y_1, y_2, \ldots, y_{m+1})^\top \in \mathbb{R}^{m+1}$ and $y^* = (y_2, \ldots, y_{m+1})^\top \in \mathbb{R}^m$, then

$$y_1 w^\top S_0 = w^\top M y^* \tag{3.1.4}$$

for all $w \in \mathbb{R}^n$, which implies $x = y^*/y_1 > 0$ is a state-price vector. $\square$

If the market is arbitrage free and $x = (x_1, \ldots x_m)^\top$ is a state-price vector, then by denoting

$$\beta = \sum_{i=1}^{m} x_i \tag{3.1.5}$$

and $\mathbb{Q}(\omega_i) = q_i = x_i/\beta > 0$ for all $i = 1, 2, \ldots, m$, we have a new probability measure $\mathbb{Q}$ on Ω. This measure has the property

$$S_0 = Mx = \beta M \frac{x}{\beta} = \beta \, \mathrm{E}_{\mathbb{Q}}(S_1). \tag{3.1.6}$$

Since we assumed that $\mathbb{P}(\omega) > 0$ for all $\omega \in \Omega$, it is easy to see that $\mathbb{Q}(\omega) > 0$ for all $\omega \in \Omega$.

A deflator d is a strictly positive process, so that $d_0 > 0$ and $d_1(\omega) > 0$ for all $\omega \in \Omega$. Now we may define relative price processes

$$S_0^d = \frac{S_0}{d_0}, \tag{3.1.7}$$

$$S_1^d = \frac{S_1}{d_1} \tag{3.1.8}$$

with regards to the deflator d.

Lemma 3.1.3. *The market has a state-space vector if and only if the market with relative prices has a state-price vector.*

Proof. Let

$$M^d = \left[\frac{S_1(\omega_1)}{d_1(\omega_1)} \; \frac{S_1(\omega_2)}{d_1(\omega_2)} \; \cdots \; \frac{S_1(\omega_m)}{d_1(\omega_m)} \right]_{n\times m}. \tag{3.1.9}$$

If $x = (x_1, x_2, \ldots, x_m)^\top > 0$ and

$$x^d = d_0^{-1}(x_1 d_1(\omega_1), x_2 d_1(\omega_2), \ldots, x_m d_1(\omega_m))^\top, \tag{3.1.10}$$

then $d_0 M^d * x^d = Mx$. Therefore $S_0 = Mx$ if and only if $S_0^d = M^d x^d$, where $x^d \in \mathbb{R}^m$ and $x^d > 0$. $\square$

Inspired by these observations, we define that a measure $\mathbb{Q}$ that satisfies

1. $\mathbb{P}(\omega) = 0$ if and only if $\mathbb{Q}(\omega) = 0$ (meaning that the measures share null sets) and

2. there exists a deflator d such that

$$S_0^d = \mathrm{E}_{\mathbb{Q}}\left(S_1^d\right) \tag{3.1.11}$$

is an equivalent martingale measure induced by deflator d. By Lemma 3.1.3, the original market is arbitrage-free if and only if the deflated market is arbitrage-free, which is equivalent to

$$S_0^d = \beta\, \mathrm{E}_{\mathbb{Q}}\left(S_1^d\right) \tag{3.1.12}$$

for some $\beta \in \mathbb{R}_+$ and measure $\mathbb{Q}$. We can define a new deflator e by $e_0 = d_0$ and $e_1 = \beta d_1$. Now

$$S_0^e = \beta\, \mathrm{E}_{\mathbb{Q}}(S_1^e) \tag{3.1.13}$$

meaning that $\mathbb{Q}$ is an EMM induced by e. Thus we see that the market is arbitrage-free if and only if it has an EMM.

The curious aspect of the probabilities of the EMM is that they are not influenced by the original probability measure $\mathbb{P}$ beyond the fact that they share the same sets of non-zero probabilities. In fact, the probabilities of the EMM is defined by the original price vector S_0 and the state space M.

Assume that $\mathbb{Q}$ is an equivalent measure. Since $\mathbb{P}$ is probability measure with no null sets, we may define a new random variable called Radon-Nikodým derivative as

$$\Lambda(\omega) = \frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)} \tag{3.1.14}$$

for all $\omega \in \Omega$. If X is a random variable, then

$$\mathrm{E}_{\mathbb{P}}(\Lambda X) = \sum_{i=1}^{m} \mathbb{P}(\omega_i)\frac{\mathbb{Q}(\omega_i)}{\mathbb{P}(\omega_i)} X(\omega_i) = \mathrm{E}_{\mathbb{Q}}(X) \tag{3.1.15}$$

so we see that the expectations under different measures are linked by the Radon-Nikodým derivative. By this we see that X is a martingale under the measure $\mathbb{Q}$ if and only if ΛX is a martingale under the measure $\mathbb{P}$.

A numéraire is any asset with only positive prices. If one of the assets is a numéraire with initial price s_0 and terminal price $s_1(\omega)$, then we use it as a deflator and define relative price process

$$S_0^* = \frac{S_0}{s_0}, \tag{3.1.16}$$

$$S_1^* = \frac{S_1}{s_1}. \tag{3.1.17}$$

If the market is arbitrage-free, then by Lemma 3.1.3 there exists such $\beta \in \mathbb{R}_+$ and measure $\mathbb{Q}$ that

$$S_0^* = \beta\, \mathrm{E}_{\mathbb{Q}}(S_1^*)\,. \tag{3.1.18}$$

But since s is one of the assets, it must hold that $\beta = 1$ meaning that the measure $\mathbb{Q}$ is an equivalent martingale measure.

We can now state the first fundamental theorem of asset pricing that weaves these different concepts together.

Theorem 3.1.4 (First fundamental theorem of asset pricing for discrete one period model)**.** *The following are equivalent in a discrete one period market model:*

1. *the market is arbitrage free,*
2. *a state-price vector exists and*
3. *an equivalent martingale measure $\mathbb{Q}$ exists.*

If $\mathbb{Q}_d$ is an EMM induced by a deflator d and $\mathbb{Q}_e$ is an EMM induced by a deflator e, *then*

$$S_0 = d_0 \operatorname{E}_{\mathbb{Q}_d}\left(S_1^d\right) = e_0 \operatorname{E}_{\mathbb{Q}_d}\left(S_1^e\right). \tag{3.1.19}$$

In particularly, if the market has a numéraire s and there is no arbitrage, then a martingale measure induced by s satisfies

$$S_0 = s_0 \operatorname{E}_{\mathbb{Q}}\left(\frac{S_1}{s_1}\right). \tag{3.1.20}$$

A risk-free asset is a numéraire with a constant terminal price z. We assume that the risk-free asset has initial price of 1 and terminal price of $1+r$, where $r > -1$. Thus if a risk-free asset exists and there is no arbitrage, an EMM $\mathbb{Q}$ induced by a risk-free asset satisfies

$$S_0 = \operatorname{E}_{\mathbb{Q}}\left(\frac{S_1}{1+r}\right). \tag{3.1.21}$$

3.1.2. The second fundamental theorem of asset pricing in discrete one period model

The arguments presented in this section is essentially the same as the ones in Björk (2004, pp. 31–34)

How should we then price derivatives in this model? A contingent claim X is a random variable $X : \Omega \to \mathbb{R}$. If the the original market is arbitrage-free, then the first fundamental theorem of asset pricing implies that an equivalent martingale measure $\mathbb{Q}$ induced by a deflator d satisfies

$$S_0 = d_0 \operatorname{E}_{\mathbb{Q}}\left(S_1^d\right). \tag{3.1.22}$$

It would be natural to define

$$X_0 = d_0 \operatorname{E}_{\mathbb{Q}}\left(X_1^d\right) \tag{3.1.23}$$

to be the initial price of the claim X. But since there may be different EMMs, X_0 may not be well-defined. Thus it is vital to pose the question how many measures there are. If the market is arbitrage free, then the equation $S_0 = Mx$ has a solution by Lemma 3.1.2. By basic linear algebra, this solution is unique if and only if $\operatorname{Ker}(M) = 0$. We also know that null space is the orthogonal compliment of the row space meaning that

$$\operatorname{Ker}(M) = \operatorname{Im}(M^\top)^\perp \tag{3.1.24}$$

and this suggests that we have a closer look at the image set $\{x^\top M \mid x \in \mathbb{R}^n\}$. But the image set is just the set of all possible portfolios of the original market.

If there is such portfolio w that $w^\top S_1 = X$ $\mathbb{P}$-surely, then we say that X is replicated by the portfolio w. This is equivalent to

$$X \in \{x^\top M \mid x \in \mathbb{R}^n\}. \tag{3.1.25}$$

By the first fundamental theorem of asset pricing, we see that if the contingent claim X can be replicated and the market has no arbitrage, then for any pair of an EMM $\mathbb{Q}$ and a replicating portfolio w, we have that

$$V_0^w = d_0 \, \mathrm{E}_{\mathbb{Q}} \left(w^\top S_1^d \right) = d_0 \, \mathrm{E}_{\mathbb{Q}} \left(\frac{X}{d} \right), \tag{3.1.26}$$

where d is a deflator inducing $\mathbb{Q}$. The left hand side depends only on the replicating portfolio w and the right hand size depends only on the measure $\mathbb{Q}$ and the deflator. This implies that $V_0^{w_1} = V_0^{w_2}$ for any portfolios w_1, w_2 which replicates X.

We define $Z_0 = (w^\top S_0, S_0^\top)^\top$, $Z_1 = (w^\top S_1, S_1^\top)^\top$. If $\mathbb{Q}$ is an EMM with a deflator d, then

$$Z_0 = d_0 \, \mathrm{E}_{\mathbb{Q}} \left(Z_1^d \right) \tag{3.1.27}$$

if and only if

$$w^\top S_0 = d_0 \, \mathrm{E}_{\mathbb{Q}} \left(w^\top S_1^d \right), \tag{3.1.28}$$

$$S_0 = d_0 \, \mathrm{E}_{\mathbb{Q}} \left(S_1^d \right). \tag{3.1.29}$$

Hence, if the market is arbitrage-free, then the arbitrage-free price of a replicated contingent claim X is $w^\top S_0$, where w is any replicating portfolio.

Now we define that the market is complete if every contingent claim can be replicated. This means that if M is defined as in 3.1.1, then the marker is complete if and only if

$$\mathrm{Im}\,(M^\top) = \{M^\top w \mid w \in \mathbb{R}^n\} = \{(w^\top M)^\top \mid w \in \mathbb{R}^n\} = \mathbb{R}^m \tag{3.1.30}$$

meaning that the matrix M has a rank of m. By duality in 3.1.24, we see that this is equivalent to $\mathrm{Ker}\,(M) = 0$. Therefore the market is complete if and only if the solution to the equation in Lemma 3.1.2 has a unique solution. Thus we have proved the second fundamental theorem of asset pricing.

Theorem 3.1.5 (Second fundamental theorem of asset pricing)**.** *Assume that the market is arbitrage free. The market is complete if and only if there is a deflator that induces a unique equivalent martingale measure. Then every EMM induced by a given deflator is unique. If the market has a numéraire, then the market is complete if and only if EMM induced by the numéraire is unique.*

Completeness also implies that the market has m linearly independent asset price processes since the matrix M has a rank of m. This means, in a sense, that for every risk dimension is covered be tradable assets and therefore every possible contingent claim can be replicated.

So we have identified three distinct scenarios which are from best to worst:

1. The market model is arbitrage free and complete which is equivalent to the fact the there exists a deflator with a unique martingale measure. Every contingent claim can be given a unique price which is the cost of replicating portfolio. Or, if d is deflator with induced EMM $\mathbb{Q}$, then the initial arbitrage-free price of a contingent claim X is

$$X_0 = d_0 \, \mathrm{E}_{\mathbb{Q}} \left(\frac{X_1}{d_1} \right). \tag{3.1.31}$$

   This price does not depend on the choice of the deflator d as long as the deflator induces an EMM.

2. The market is arbitrage free but not complete. Then every deflator which induces an equivalent martingale measure have several EMMs. Every replicated contingent claim can be given a unique arbitrage-free price which is the cost of replicating portfolio. This price is given by Equation 3.1.31. Contingent claims that could not be replicated may not be priced in the sense of the Equation 3.1.31.

3. The market has arbitrage which makes pricing rather meaningless.

Thus, if the market does not allow arbitrage, then we may price every replicated contingent claim can be given an arbitrage-free price. This price is the cost of the cost of replicating portfolio w which is equal to

$$V_0^w = d_0 \, \mathrm{E}_{\mathbb{Q}} \left(\frac{V_1^w}{d_1} \right), \tag{3.1.32}$$

where d is any deflator that induces an EMM $\mathbb{Q}$. If the market has a numéraire, then we may use this as the deflator.

3.2. Arbitrage theory in continuous markets

We now develop an heuristic model for continuous-time markets. We shall omit most technical definitions, details and proofs. It is an amalgam of the presentations found in Björk (2004), Brigo et al. (2007), Duffie (2010), James and Webber (2000), Musiela et al. (2005) and Wu (2009). A rigorous treatment for the subject can be found, for example, in Musiela et al. (2005). The fundamentals of this model are the same as the simple discrete model introduced earlier, but due to technicalities, it is not as intuitive.

We consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a finite time interval $[0, T']$. Here $\mathbb{P}$ is the physical probability measure on $(\Omega, \mathscr{F})$, Ω is the sample path space with a σ-algebra $\mathscr{F}$. The flow of new information is handled with a filtration[1] $(\mathscr{F}_t)$. We also assume some technical conditions for the filtrations. Every $\mathbb{P}$-null set[2] must be a member of $\mathscr{F}_0$ and

$$\mathscr{F}_t = \bigcap_{t<s} \mathscr{F}_s \tag{3.2.1}$$

for all $t \geq 0$.

We assume that the price S_i of a market assets are modeled with Itó-processes[3], so

$$S_i(t,\omega) = S_i(0) + \int_0^t \mu(t,\omega)\mathrm{d}t + \int_0^t \sigma(t,\omega)\mathrm{d}W_i(t,\omega), \tag{3.2.2}$$

where $S_i(0)$ is a deterministic constant and W_i is a Brownian motion. The second integral is an Itó-integral and the functions μ and σ are assumed to be $\mathscr{F}_t$-adapted and to satisfy technical conditions so that integrability and the existence of solutions are always guaranteed[4]. We usually write the Equation 3.2.2 as

$$\mathrm{d}S_i(t,\omega) = \mu(t,\omega)\mathrm{d}t + \sigma(t,\omega)\mathrm{d}W_i(t,\omega), \tag{3.2.3}$$

but we note that Brownian motions are $\mathbb{P}$-surely not differentiable. We assume that the assets all these assets can be bought and sold freely, and the trading shall not affect price process.

We also shall omit ω from the argument of functions for readability, when it is not absolutely necessary.

Suppose that there are n tradable assets and let $S = (S_1, S_2, \ldots, S_n)$ be the price vector for these assets. The trading strategy (or a portfolio) w is a predictable stochastic

[1] A filtration $(\mathscr{F}_t)$ is a collection σ-sub-algebras of $\mathscr{F}$ with $\mathscr{F}_s \subseteq \mathscr{F}_t$ for all $0 \leq s \leq t \leq T$.

[2] $\mathbb{P}$-null set is a set $A \in \mathscr{F}$ with $\mathbb{P}(A) = 0$.

[3] See A.8

[4] It is usually assumed, for example, that $\int_0^{T^*} |\mu(t,\omega)|\,\mathrm{d}t < \infty$ and $\int_0^{T^*} \sigma(t,\omega)^2\mathrm{d}t < \infty$ almost surely

process $w(t,\omega) \in \mathbb{R}^n$. This means that while the strategy is a random variable, it is not omniscient and uses only information available up to that point. It is left-continuous, so given the history up to that point, it is deterministic. The corresponding portfolio value process is given by

$$V^w(t) = (w(t))^\top S(t) \tag{3.2.4}$$

and a self-financing trading strategy satisfies

$$V^w(t) - V^w(0) = \int_0^t w^\top(u)\, \mathrm{d}S(u), \tag{3.2.5}$$

or

$$\mathrm{d}V^w = (w(t))^\top \mathrm{d}S(t). \tag{3.2.6}$$

This means that for self-financing portfolio, there is no inflow or outflow of money from the portfolio and all the value changes are due to changes of prices. An arbitrage[5] is a self-financing portfolio with a value process $V^w(t)$ such that

$$V^w(0) = 0, \tag{3.2.7}$$
$$\mathbb{P}(V^w(t) \geq 0) = 1 \text{ and} \tag{3.2.8}$$
$$\mathbb{P}(V^w(t) > 0) > 0. \tag{3.2.9}$$

for some $0 \leq t \leq T'$. An arbitrage is an opportunity to gain without any associated risk. A market without arbitrage opportunities is arbitrage-free.

A contingent claim (or T-derivative or T-claim) is a $\mathscr{F}_T$-measurable random variable X. We say that the is attainable if there is such self-financing trading strategy w that

$$V^w(T) = X(T) \tag{3.2.10}$$

almost surely. The strategy w is thus replicating the pay-off of X. As usually, we say that the market is complete if every contingent claim is attainable.

As earlier, a deflator is strictly positive Itô-process and a numéraire is an asset with always positive price process. If D is deflator, then the relative (or discounted) market

[5]We note that this type of arbitrage is not strong enough to disallow doubling strategies.

price process is

$$S^D(t) = (S_i(t)/D(t)) \tag{3.2.11}$$

for all $0 \leq t \leq T'$. A measure $\mathbb{Q}$ on the space $(\Omega, \mathscr{F})$ is an equivalent martingale measure (EMM) with respect to the deflator D if it is equivalent[6] to $\mathbb{P}$ and if

$$S^D(s) = \mathrm{E}_{\mathbb{Q}}(\ S^D(t) \mid \mathscr{F}_s\ ) \tag{3.2.12}$$

for all $0 \leq s < t$. This means that discounted price processes are martingales under an EMM. A weaker condition is an equivalent local martingale measure (ELMM) which requires only that the measures are equivalent and discounted price process are local martingale under the ELMM.

We will only consider trading strategies that will satisfy the condition that

$$\int_0^t w(s)^\top \mathrm{d}S^D(s) \tag{3.2.13}$$

are martingales under the the measure $\mathbb{Q}$ as this guarantees that no doubling strategies are permissable[7].

The first fundamental theorem of asset pricing says that if an equivalent martingale measure exists, then the market is arbitrage-free. As we have seen, in discrete time model, these are equivalent conditions. But in continuous models, this in not true. Delbaen and Schachermayer (1994) showed that under certain conditions, the absence of arbitrage implies that an equivalent local martingale measure exists. We shall not delve in these technicalities further in this thesis.

The second fundamental theorem of asset pricing ties the uniqueness of this EMM to the ability to hedge every derivative contract. If we assume that the market is arbitrage free and an equivalent martingale measure exists for a deflator D, then the market is complete if and only if the equivalent martingale measure for deflator D is unique.

As in discrete setting, we have three distinct cases: the model has no EMM, it has several EMMs or the EMM is unique.

Ideally the EMM is unique and the model is complete and it allows no arbitrage. Traditionally the discounting in the EMM has been done with respect to the risk-free rate, but Geman, El Karoui, and Rochet (1995) showed that the choice of the so-called numéraire is actually arbitrary as long as it is strictly positive non-dividend paying asset process. Different numéraires produces different EMMs, but if we assume that payoffs

[6] Meaning that the measures have the same null sets.
[7] Thus the expected values of portfolios are always bounded.

are square integrable random variables, then the change of numéraire does not change replicating portfolios. Thus the price is unique as long as the derivative can be hedged. If S_0 is the numéraire, X is a T-derivate with replicating self-financing portfolio's value process given by $V^w(t)$, then we have that $X(T) = V^w(T)$ and every discounted value processes of self-financing portfolio is martingale under the EMM $\mathbb{Q}_0$ associated with discounting process S_0. Thus

$$\frac{V^w(t)}{S_0(t)} = \mathrm{E}_{\mathbb{Q}_0}\left(\frac{V^w(T)}{S_0(T)}|\mathscr{F}_t\right) = \mathrm{E}_{\mathbb{Q}_0}\left(\frac{X(T)}{S_0(T)}|\mathscr{F}_t\right) \tag{3.2.14}$$

meaning that

$$V^w(t) = S_0(t)\,\mathrm{E}_{\mathbb{Q}_0}\left(\frac{V^w(T)}{S_0(T)}|\mathscr{F}_t\right) = S_0(t)\,\mathrm{E}_{\mathbb{Q}_0}\left(\frac{X(T)}{S_0(T)}|\mathscr{F}_t\right). \tag{3.2.15}$$

Since there are no arbitrage opportunities, $X(t) = V^w(t)$, where $V^w(t)$ is given by the equation (3.2.15). This has to hold even if the replicating portfolio is not unique.

If the model has several EMMs, then arbitrage is not possible but there are derivatives that may not be hedged. In this model, a claim that can be replicated has a unique price. Claims that may not be replicated may have multiple prices corresponding to different EMMs.

This usually means that calibrating prices are taken carefully selected from the most liquid instruments. We can also use more generalized versions of hedging. We could take the set of all sensible portfolio strategies that promise almost surely a payoff that is equal or greater than the contingent claim. A reasonable price candidate for a derivative is the minimum maintenance price of these portfolio strategies.

The worst case is the absence of EMM. This means that the model has arbitrage and pricing cannot be done.

The numéraire can be freely chosen. Geman et al. 1995 showed that if the market is arbitrage-free, M and N are arbitrary numéraires, then there exists such EMMs $\mathbb{Q}_M$ and $\mathbb{Q}_n$ that

$$\frac{X(t)}{M(t)} = \mathrm{E}_{\mathbb{Q}_M}\left(\frac{X(T)}{M(T)}|\mathscr{F}_t\right), \tag{3.2.16}$$

$$\frac{X(t)}{N(t)} = \mathrm{E}_{\mathbb{Q}_N}\left(\frac{X(T)}{N(T)}|\mathscr{F}_t\right) \tag{3.2.17}$$

for any asset X and $0 \leq t \leq T$. The Radon-Nikodým derivate[8] is

$$\xi(t) = \frac{\mathrm{d}\mathbb{Q}_M}{\mathrm{d}\mathbb{Q}_N} = \frac{M(t)N(0)}{M(0)N(t)}. \tag{3.2.18}$$

This implies that

$$X(t) = M(t)\,\mathrm{E}_{\mathbb{Q}_M}\left(\frac{X(T)}{M(T)}|\mathscr{F}_t\right) \tag{3.2.19}$$

$$= N(t)\,\mathrm{E}_{\mathbb{Q}_N}\left(\frac{X(T)}{N(T)}|\mathscr{F}_t\right) \tag{3.2.20}$$

and the price is unique, if the claim can be replicated. Also

$$\mathrm{E}_{\mathbb{Q}_M}\left(Y(T)|\mathscr{F}_t\right) = \frac{\mathrm{E}_{\mathbb{Q}_N}\left(Y(T)\xi(T)|\mathscr{F}_t\right)}{\xi(t)} \tag{3.2.21}$$

holds for any random variable Y.

3.2.1. Risk-free measure

The bank account $B(t) > 0$ is a common numéraire and the EMM $\mathbb{Q}_0$ induced by it is often called as the risk-free measure. If X is a portfolio and the market is arbitrage-free, then

$$X(t) = B(t)\,\mathrm{E}_{\mathbb{Q}_0}\left(\frac{X(T)}{B(T)}|\mathscr{F}_t\right) \tag{3.2.22}$$

$$= \mathrm{E}_{\mathbb{Q}_0}\left(\frac{B(t)}{B(T)}X(T)|\mathscr{F}_t\right) \tag{3.2.23}$$

$$= \mathrm{E}_{\mathbb{Q}_0}\left(D(t,T)X(T)|\mathscr{F}_t\right), \tag{3.2.24}$$

where $D(t,T)$ is the stochastic discount factor. If $X(t) = p(t,T)$, then $p(T,T) = 1$ and we note that

$$p(t,T) = \mathrm{E}_{\mathbb{Q}_0}\left(D(t,T) \mid \mathscr{F}_t\right). \tag{3.2.25}$$

This shows that price of a bond is expected value of the corresponding stochastic discount factor under the risk-free measure (or any other EMM). Also if

$$B(t) = \mathrm{e}^{\int_0^t r(s)\,\mathrm{d}s}, \tag{3.2.26}$$

[8] See A.2

then

$$D(t,T) = \frac{B(t)}{B(T)} = \mathrm{e}^{-\int_t^T r(s)\ \mathrm{d}s}, \tag{3.2.27}$$

hence

$$X(t) = \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^T r(s)\ \mathrm{d}s} X(T)|\mathscr{F}_t\right). \tag{3.2.28}$$

Under the risk-neutral measure, the discounted process $X(t)/B(t)$ will be a martingale.

3.2.2. Black-Scholes–model

In the celebrated Black-Scholes–model there are two assets, a stock and a bank account, with given dynamics

$$\mathrm{d}S(t) = \mu S(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W(t) \tag{3.2.29}$$

$$\mathrm{d}B(t) = rB(t)\mathrm{d}t, \tag{3.2.30}$$

under the physical measure $\mathbb{P}$, where μ, r and $\sigma > 0$ are given constants and the Brownian motion $W(t)$ is the sole source of uncertainty. This means that the stock price follows geometric Brownian motion and $B(t) = B(0)\mathrm{e}^{rt}$. To simplify the notation we assume that $B(0) = 1$. We denote the discounted stock price as $S^B(t) = S(t)/B(t)$ and the discount factor is deterministic $D(t,T) = \mathrm{e}^{r(T-t)}$. Now $g(t,x) = \mathrm{e}^{-rt}x$ and a simple application of Itô's lemma[9] yields that

$$\mathrm{d}S^B(t) = (\frac{\partial g}{\partial t} + \mu\frac{\partial g}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 g}{\partial x^2})\mathrm{d}t + \sigma\frac{\partial g}{\partial x}\mathrm{d}W(t) \tag{3.2.31}$$

$$= (-r\mathrm{e}^{-rt}S(t) + \mathrm{e}^{-rt}\mu S(t) + 0)\mathrm{d}t + \sigma\mathrm{e}^{-rt}S(t)\mathrm{d}W(t) \tag{3.2.32}$$

$$= (\mu - r)S^B(t)\mathrm{d}t + \sigma S^B(t)\mathrm{d}W(t). \tag{3.2.33}$$

We can now use Girsanov's theorem[10] to change the probability measure in order to make the discounted stock price process driftless. We note the bank account discounted by itself is trivially driftless under any measure. Now the market price of risk

$$\lambda = \frac{r-\mu}{\sigma} \tag{3.2.34}$$

[9] See A.8.
[10] See A.10.

is the only possible Girsanov kernel that makes the new measure an equivalent martingale measure. So there is an unique EMM $\mathbb{Q}$ for the deflator $B(t)$. hence Black-Scholes model is arbitrage-free and complete. Under this measure $\mathrm{d}W^{\mathbb{Q}}(t) = \mathrm{d}W(t) - \frac{r-\mu}{\sigma}\mathrm{d}t$ is a Brownian motion. Therefore

$$\mathrm{d}S^B(t) = \sigma S^B(t)\mathrm{d}W^{\mathbb{Q}}(t), \tag{3.2.35}$$

$$\mathrm{d}B(t) = rB(t)\mathrm{d}t, \tag{3.2.36}$$

$$\mathrm{d}S(t) = rS(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W^{\mathbb{Q}}(t) \tag{3.2.37}$$

under the EMM $\mathbb{Q}$. In other words, the discounted stock price is a martingale and under this measure the drift of the stock price is changed to the risk-free rate r. Since the bank account is deterministic, the change of measure does not affect it. The arbitrage-free price of the derivate is at time 0 is given by

$$X(0) = \mathrm{E}_{\mathbb{Q}}\left(X(T)/B(0,T)|\mathscr{F}_0\right) \tag{3.2.38}$$

$$= \mathrm{e}^{-rT}\,\mathrm{E}_{\mathbb{Q}}\left(X(T)|\mathscr{F}_0\right). \tag{3.2.39}$$

Since the stock price process under the risk-free measure is a geometric Brownian motion, we know that

$$\log(S(T)) = \log(S(0)) + \left(r - \frac{1}{2}\sigma^2\right)T + \sigma W^{\mathbb{Q}}(T) \tag{3.2.40}$$

$$\sim N\left(\log(S(0)) + \left(r - \frac{1}{2}\sigma^2\right), \sigma^2 T\right) \tag{3.2.41}$$

under the measure $\mathbb{Q}$. Similarly, we could use the stock price $S(t)$ as the numerator.

If we pick $X(T) = (S(T)-K)^+$, the price of a call option on the stock at the time T, then

$$\mathrm{E}_{\mathbb{Q}}\left((S(T)-K)^+|\mathscr{F}_0\right) = \mathrm{E}_{\mathbb{Q}}\left((S(T)-K)1_{\{S(T)>K\}}|\mathscr{F}_0\right) \tag{3.2.42}$$

$$= \mathrm{E}_{\mathbb{Q}}\left(S(T)1_{\{S(T)>K\}}|\mathscr{F}_0\right) - K\mathbb{Q}\left(S(T)>K|\mathscr{F}_0\right) \tag{3.2.43}$$

and the standard calculations will yield that

$$X(T) = S(0)N(d_+) - \mathrm{e}^{-rT}KN(d_-), \tag{3.2.44}$$

$$d_\pm = \frac{\log S/K + (r \pm \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{3.2.45}$$

which is the celebrated Black-Scholes formula.

We shall also give a heuristic derivation of the Black-Scholes differential equation. If $X(t,S(t))$ is a smooth value process of a derivative, then by Itó's lemma, we have that

$$\mathrm{d}X = \left(\frac{\partial X}{\partial t} + \mu S\frac{\partial X}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial X^2}{\partial S^2}\right)\mathrm{d}t + \sigma S\frac{\partial X}{\partial S}\mathrm{d}W. \tag{3.2.46}$$

Now we assume that we can replicate this derivative with a combination of $\delta_S(t,S)$ stocks and $\delta_B(t,S)$ bonds. This portfolio has a value

$$V(t,S) = \delta_S S(t) + \delta_B B(t) \tag{3.2.47}$$

and the value follows the process

$$\begin{aligned} \mathrm{d}V &= \delta_S \mathrm{d}S + \delta_B \mathrm{d}B && (3.2.48)\\ &= \delta_S(\mu S\mathrm{d}t + \sigma S\mathrm{d}W) + \delta_B rB\mathrm{d}t && (3.2.49)\\ &= (\delta_S \mu S + \delta_B rB)\mathrm{d}t + \delta_S \sigma S\mathrm{d}W && (3.2.50)\\ &= (\delta_S \mu S + r(V - \delta_S S))\mathrm{d}t + \delta_S \sigma S\mathrm{d}W && (3.2.51) \end{aligned}$$

As $\mathrm{d}V = \mathrm{d}X$, then the coefficients of $\mathrm{d}W$ terms must coincide. Thus

$$\delta_S \sigma S = \sigma S\frac{\partial X}{\partial S}, \tag{3.2.52}$$

which implies that $\delta_S = \frac{\partial X}{\partial S}$.

Now we consider a portfolio of one derivative and $-\frac{\partial X}{\partial S}$ stocks. The value of this portfolio is

$$V = X - \frac{\partial X}{\partial S}S \tag{3.2.53}$$

and if the portfolio is self-financing, then

$$\begin{aligned} \mathrm{d}V &= \mathrm{d}X - \frac{\partial X}{\partial S}\mathrm{d}S && (3.2.54)\\ &= \left(\frac{\partial X}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial X^2}{\partial S^2}\right)\mathrm{d}t && (3.2.55) \end{aligned}$$

after substitution of equations 3.2.29 and 3.2.46. Since there is no diffusion, the portfolio is riskless and the absence of arbitrage implies that

$$\mathrm{d}V = rV\mathrm{d}t = \left(rX - rS\frac{\partial X}{\partial S}\right)\mathrm{d}t. \tag{3.2.56}$$

By equating these, we get the Black-Scholes differential equation

$$\frac{\partial X}{\partial t} + rS\frac{\partial X}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial X^2}{\partial S^2} - rX = 0 \tag{3.2.57}$$

or equivalently

$$\frac{\partial X}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial X^2}{\partial S^2} = r(X - S\frac{\partial X}{\partial S}). \tag{3.2.58}$$

The left side is a linear combination of "theta", the time decay of the value, and "gamma", the second derivative of the value with respect to the price of the underlying. The right side of the equation contains the replicating portfolio.

The derivative with pay-out $h(S(T))$ satisfies Equation 3.2.57. In order to apply Feynman-Kac theorem[11], we need to find a process with drift $rS(t)$ and diffusion $\sigma S(t)$ under some measure. Under the EMM, the discounted asset price is a martingale, so it grows with the rate risk-free rate r. So the price process under the EMM is process we need in order to use Feynman-Kac theorem. We have the EMM $\mathbb{Q}$ and, by Equation 3.2.37,

$$\mathrm{d}S(t) = rS(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W^{\mathbb{Q}}(t), \tag{3.2.59}$$

where $W^{\mathbb{Q}}(t)$ is a Brownian motion under $\mathbb{Q}$. We may now apply Feynman-Kac theorem and it follows that

$$h(S(t)) = \mathrm{e}^{-r(T-t)}\,\mathrm{E}_{\mathbb{Q}}(h(S_T) \mid \mathscr{F}_t). \tag{3.2.60}$$

3.2.3. Black-76–model

Black model (or Black-76–model) is an extension of Black-Scholes–model (Black 1976) and it is used to price futures. It also assumes that the risk-free interest rate is a constant. The model assumes that futures price of an asset follows log-normal distribution with constant volatility parameter. The price of a call option on a future

[11] See A.12.

contract has price at time t is given by the Black's formula

$$e^{-r(T-t)}\left(FN(d_+) - KN(d_-)\right) \tag{3.2.61}$$

with

$$d_\pm = \frac{\log\frac{F}{K} \pm \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{(T-t)}}. \tag{3.2.62}$$

Here K is the strike price at the maturity T, r is the risk-free rate and σ is the constant volatility of the log-normal distribution. The futures price process is $F(t)$. The Black's formula is used to price interest rate caps, floors and swaptions and the market practice is to quite these instruments in terms of Black's volatilities.

3.2.4. T-forward measure

Since $p(t,T) > 0$, T-bond is a numéraire. The EMM induced by this as called T-forward measure $\mathbb{Q}_T$. Since $p(T,T) = 1$, we have that

$$X(t) = p(t,T)\,\mathrm{E}_{\mathbb{Q}_T}\left(X(T)|\mathscr{F}_t\right) \tag{3.2.63}$$

for every $0 \le t \le T$ and attainable claim X. The forward rate was defined as

$$L(t,T,S) = \frac{1}{\Delta(T,S)}\left(\frac{p(t,T)}{p(t,S)} - 1\right). \tag{3.2.64}$$

Thus

$$L(t,T,S)p(t,S) = \frac{p(t,T) - p(t,S)}{\Delta(T,S)}, \tag{3.2.65}$$

where the right side is a bond portfolio. As $p(T,T) = 1$, we know that

$$\frac{p(t,T) - p(t,S)}{\Delta(T,S)} = p(t,T)\,\mathrm{E}_{\mathbb{Q}_T}\left(L(T,T,S)|\mathscr{F}_t\right) \tag{3.2.66}$$

Now we have that shown that

$$L(t,T,S) = \mathrm{E}_{\mathbb{Q}_T}\left(L(T,S)|\mathscr{F}_t\right) \tag{3.2.67}$$

meaning that the forward rate is expected value of spot rate under the T-forward measure.

Similarly

$$P(t,T)\,\mathrm{E}_{\mathbb{Q}_T}\left(r(T)|\mathscr{F}_t\right) = \mathrm{E}_{\mathbb{Q}_0}\left(D(t,T)r(T)|\mathscr{F}_t\right) \tag{3.2.68}$$

$$= \mathrm{E}_{\mathbb{Q}_0}\left(r(T)\mathrm{e}^{-\int_t^T r(s)\mathrm{d}s}|\mathscr{F}_t\right) \tag{3.2.69}$$

$$= \mathrm{E}_{\mathbb{Q}_0}\left(\frac{\partial}{\partial T}\mathrm{e}^{-\int_t^T r(s)}|\mathscr{F}_t\right) \tag{3.2.70}$$

$$= \frac{\partial}{\partial T}\mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^T r(s)}|\mathscr{F}_t\right) \tag{3.2.71}$$

$$= \frac{\partial}{\partial T}P(t,T), \tag{3.2.72}$$

which implies that

$$f(t,T) = \mathrm{E}_{\mathbb{Q}_T}\left(r(T)|\mathscr{F}_t\right) \tag{3.2.73}$$

meaning that the instantaneous forward rate is the expected value of the short-rate under T-forward measure.

The Equation 3.2.21 implies that

$$\xi(t) = \frac{p(t,T)}{p(0,T)B(t)} \tag{3.2.74}$$

is the Radon-Nikodým-derivative of T-forward measure $\mathbb{Q}_T$ with respect to risk-free measure $\mathbb{Q}_0$.

3.2.5. Change of numéraire

Now we consider an arbitrage free market model with assets N and M, which are numéraires. If X is a contingent T-claim, then we know that arbitrage free price of X at the time t must be

$$V_t(X) = N(t)\,\mathrm{E}_{\mathbb{Q}_N}\left(\frac{X(T)}{N(T)}|\mathscr{F}_t\right) \tag{3.2.75}$$

$$= M(t)\,\mathrm{E}_{\mathbb{Q}_M}\left(\frac{X(T)}{M(T)}|\mathscr{F}_t\right) \tag{3.2.76}$$

where $\mathbb{Q}_N, \mathbb{Q}_M$ are the martingale measure corresponding to the numéraires N and M. Thus

$$\mathrm{E}_{\mathbb{Q}_N}\left(\frac{X(T)}{N(T)}|\mathscr{F}_t\right) = \mathrm{E}_{\mathbb{Q}_M}\left(\frac{M(t)}{N(t)}\frac{X(T)}{M(T)}|\mathscr{F}_t\right) \tag{3.2.77}$$

$$= \mathrm{E}_{\mathbb{Q}_M}\left(L_T(t)\frac{X(T)}{N(T)}|\mathscr{F}_t\right), \tag{3.2.78}$$

where

$$L_T(t) = \frac{N(T)/N(t)}{M(T)/M(t)} = \frac{M(t)}{N(t)}\frac{N(T)}{M(T)}. \tag{3.2.79}$$

Now $L_T(t)$ is a $\mathbb{Q}_N$-martingale Since X is an arbitrary $\mathscr{F}_T$-measurable random variable, we have heuristically shown the following fundamental result. For the proof, see Geman et al. 1995.

Theorem 3.2.1. *Let $\mathbb{Q}_N$ be the EMM associated with numéraire N and $\mathbb{Q}_M$ be the EMM associated with numéraire M. Under some technical conditions, the Radon-Nikodým derivate of $\mathbb{Q}_M$ with respect to $\mathbb{Q}_N$ is*

$$\frac{\mathrm{d}\mathbb{Q}_N}{\mathrm{d}\mathbb{Q}_M} = \frac{N(T)/N(t)}{M(T)/M(t)}. \tag{3.2.80}$$

Since $Z(t,T) > 0$ for all pairs (t,T) we may use zero-coupon bonds as numéraire. We denote $\mathbb{Q}_T$ as the equivalent martingale measure given by the T-bond. Thus if X is a contingent T-claim, then we know that arbitrage free price of X at the time t must be

$$V_t(X) = Z(t,T)\,\mathrm{E}_{\mathbb{Q}_T}(X(T)). \tag{3.2.81}$$

4. SHORT-RATE MODELS

This chapter introduces basics of short-rate modeling and short-rate models iwth affine term-structures.

4.1. Introduction to short-rate models

4.1.1. Term-structure equation

The derivation of the term-structure equation follows closely Björk (2004, pp. 319–324).

In this chapter we assume that the short rate process follows

$$\mathrm{d}r(t) = \mu(t, r(t))\mathrm{d}t + \sigma(t, r(t))\mathrm{d}W(t), \tag{4.1.1}$$

where $W(t)$ is a Brownian motion under some measure $\mathbb{Q}^*$ (this may be also physical measure) and μ and σ are well-behaved functions. If $V(t, r(t))$ is a smooth function, then Itós lemma says that

$$\mathrm{d}V = \left(\frac{\partial V}{\partial t} + \mu\frac{\partial V}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial x^2}\right)\mathrm{d}t + \sigma\frac{\partial V}{\partial x}\mathrm{d}W \tag{4.1.2}$$

Here we have dropped arguments for functions. Unlike in Black-Scholes model, we may not trade short rate $r(t)$ directly and it may not be used in hedging. But we may consider how to hedge a derivative using another contingent claim.

We consider two contingent claims with value processes $V_1(t, r(t))$ and $V_2(t, r(t))$. Itó's lemma gives us that

$$\mathrm{d}V_i = M_i\mathrm{d}t + N_i\mathrm{d}W, \tag{4.1.3}$$

where

$$M_i = \frac{\partial V_i}{\partial t} + \mu\frac{\partial V_i}{\partial r} + \frac{1}{2}\sigma^2\frac{\partial^2 V_i}{\partial r^2}, \tag{4.1.4}$$

$$N_i = \sigma\frac{\partial V_i}{\partial r}. \tag{4.1.5}$$

If $\Pi = V_1 + \delta V_2$ is a portfolio, then

$$\mathrm{d}\Pi = (M_1 + \delta M_2)\,\mathrm{d}t + (N_1 + \delta N_2)\,\mathrm{d}W. \tag{4.1.6}$$

By choosing $\delta = -\frac{N_1}{N_2}$, the Brownian motion disappears. In order to maintain absence of arbitrage, we have that

$$\mathrm{d}\Pi = r\Pi\mathrm{d}t \tag{4.1.7}$$

$$\left(M_1 - \frac{N_1}{N_2}M_2\right)\mathrm{d}t = r\left(V_1 - \frac{N_1}{N_2}V_2\right)\mathrm{d}t. \tag{4.1.8}$$

Thus

$$\frac{M_1 - rV_1}{N_1} = \frac{M_2 - rV_2}{N_2} \tag{4.1.9}$$

Now M_i, N_i and V_i are functions of t and $r(t)$. The left-hand side is independent of the portfolio V_2 and the right-hand size is independent of portfolio V_1. Hence there exist a function

$$\lambda(t, r(t)) = \frac{M(t, r(t)) - r(t)V(t, r(t))}{N(t, r(t))} \tag{4.1.10}$$

called the market-price of risk, where V is any interest-rate derivative with dynamics

$$\mathrm{d}V = M\mathrm{d}t + N\mathrm{d}W. \tag{4.1.11}$$

By combining this with earlier results, we get that the price process V must satisfy partial differential equation

$$\begin{aligned}
0 &= M - rV - \lambda N && (4.1.12)\\
&= \frac{\partial V}{\partial t} + \mu\frac{\partial V}{\partial r} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r^2} - rV - \lambda\sigma\frac{\partial V}{\partial r} && (4.1.13)\\
&= \frac{\partial V}{\partial t} + (\mu - \lambda\sigma)\frac{\partial V}{\partial r} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r^2} - rV && (4.1.14)
\end{aligned}$$

with the boundary condition given by the value $V(T, r(T))$.

4.1.2. Fundamental models

Suppose that $\mathbb{Q}^* = \mathbb{P}$ is the physical measure. Models that are defined under the physical measure, are often called as fundamental models. If we denote

$$\theta(t,r(t)) = \mu(t,r(t)) - \lambda(t,r(t))\sigma(t,r(t)) \tag{4.1.15}$$

and assume that

$$\lambda(t,r(t)) = \frac{\theta(t,r(t)) - \mu(t,r(t))}{\sigma(t,r(t))} \tag{4.1.16}$$

may be used as a Girsanov kernel, then we get a new measure $\mathbb{Q}_\theta$ under which

$$r(t) = \theta(t,r(t))\mathrm{d}t + \sigma(t,r(t))\mathrm{d}W_\theta(t) \tag{4.1.17}$$

and $W_\theta(t)$ is a Brownian motion. Now Feynman-Kac theorem implies that

$$V(t) = \mathrm{E}_{\mathbb{Q}_\theta}\left(\mathrm{e}^{-\int_t^T r(s)\mathrm{d}s} V(T) \mid \mathscr{F}_t\right). \tag{4.1.18}$$

Unlike in the Black-Scholes model, the market price of risk λ is not uniquely determined endogenously within the model. Here, the market price of risk is endogenously determined. The equivalent martingale measure (if it exists), it is not unique and prices depend on the choice of the function λ. However, the λ is uniquely determined by the price of any interest rate derivative. Since the dynamic of Equation 4.1.1 are under the physical measure, we could use econometric time-series to estimate model parameters and market price of risk. This approach is problematic since neither the short-rate nor the market price of risk are not directly observable. According to Chapman, Long, and Pearson 1999, using proxies to estimate parameters of short-rate models for single-factor affine models does not cause economically significant problems, but for more complex models, proxies cause significant errors. Also, for estimation, we usually have to assume a functional form for the market price of risk, which may be misspecified. Fundamental models have no guarantee that they will fit the observed term or volatility structures. If these obstacles are overcome, then the model may be used to price any instrument and forecast interest rates.

4.1.3. Preference-free models

Another approach is to assume that the Equation 4.1.1 holds under the risk-free measure $\mathbb{Q}_0$. This means choosing $\lambda = 0$ in the Equation 4.1.12, which will now read

$$0 = \frac{\partial V}{\partial t} + \mu \frac{\partial V}{\partial r} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2} - rV. \tag{4.1.19}$$

Therefore Feynman-Kac theorem implies that

$$V(t) = \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^T r(s)\mathrm{d}s} V(T) \mid \mathscr{F}_t\right). \tag{4.1.20}$$

for all assets. Thus all discounted asset price process, where the numéraire is the bank account, are martingales under $\mathbb{Q}_0$.

Under this methodology, we may use calibrate the model parameters using observed market prices. We may take liquid instruments and then use their prices to calibrate the model. We may not use historical data in calibration, since the assumed process is under the risk-free measure. The physical measure will be different from the risk-free measure, unless we explicitly make the strong assumption that the market price of risk will be zero. However, as the volatility term does not change under measure changes, diffusion term may be estimated with data that is collected under the risk-free measure. Models under this methodology may or may not be guaranteed to fit the observed term and volatility structures. They could be useful for pricing. They may not be used to forecast prices or interest rates without further assumptions.

Preference-free +-models

Preference-free models which have constant parameters k, θ, σ are notated with single plus sign. They do not probably fit the observed term and volatility structures. Most of the early fundamental models can be also interpreted as preference-free $+$-models.

Preference-free ++-models

Preference-free models which have constant parameters k, σ but time-varying $\theta(t)$ are notated with double plus sign. They can be made to fit the observed term structure but they probably do not match the volatility surface. J. Hull et al. (1990) is often the prototypical example of a $++$-model.

When matched to term-structure and calibrated with cap or swaption prices, they can be useful in pricing of exotic interest-rate options.

Preference-free +++-models

Preference-free models which have time-varying $\theta(t), \sigma(t)$ (and sometimes k) are notated with triple plus sign. They can be made to fit both the observed term structure and some of the volatility surface. Variants of Hull-White models are sometimes $+++$-models.

Triple plus models can be useful in pricing but they could be prone to over-fitting.

4.2. One-factor short-rate models

One-factor short-rate model is model that has one underlying state variable that drives the evolution of the interest rates. This factor is often the short-rate itself. Often used form for one-factor short-rate process is

$$\mathrm{d}r(t) = k(\theta(t) - r(t))\mathrm{d}t + \sigma(t)r(t)^{\gamma}\mathrm{d}W(t) \tag{4.2.1}$$

The following table gives a quick overview of some of the models of this form.

Model	$k(t)$	$\theta(t)$	$\sigma(t)$	γ
Vašíček 1977	k	θ	σ	$-$
Dothan 1978	$-k(t)$	$-$	σ	$-$
Cox et al. 1985	k	θ	σ	$\frac{1}{2}$
J. Hull et al. 1990	k	$\theta(t)$	$\sigma(t)$	$-$

4.2.1. Affine one-factor term-structures models

The derivation of the following theorem is based on Björk (2004, pp. 329–331).

A one-factor short-rate model has affine term-structure if the zero-coupon bond price is

$$p(t,T) = \mathrm{e}^{A(t,T)-B(t,T)r(t)} \tag{4.2.2}$$

for all $0 \leq t \leq T$, where $A(t,T)$ and $B(t,T)$ deterministic and smooth functions. Since $p(T,T) = 1$, we have that

$$A(T,T) = B(T,T) = 0. \tag{4.2.3}$$

Under the assumption of affinity the instantaneous forward rate is

$$f(t,T) = -\frac{\partial \log p(t,T)}{\partial T} \tag{4.2.4}$$
$$= \frac{\partial (B(t,T)r(t) - A(t,T))}{\partial T} \tag{4.2.5}$$
$$= \frac{\partial B(t,T)}{\partial T} r(t) - \frac{\partial A(t,T)}{\partial T} \tag{4.2.6}$$

for all $0 \leq t \leq T$.

If we denote that $p(t,T) = F(t,T,r)$ to make clear that $p(t,T)$ is also a function of r, then we have that

$$\frac{\partial F}{\partial t} = \left(\frac{\partial A}{\partial t} - \frac{\partial B}{\partial t} r\right) F \tag{4.2.7}$$
$$\frac{\partial F}{\partial r} = -BF \tag{4.2.8}$$
$$\frac{\partial^2 F}{\partial r^2} = B^2 F \tag{4.2.9}$$

as the derivative of r with respect to t vanishes. Since the price of a zero-coupon price must satisfy 4.2.2, we have that

$$\frac{\partial F}{\partial t} + \mu^* \frac{\partial F}{\partial r} + \frac{1}{2}\sigma^2 \frac{\partial^2 F}{\partial r^2} - rF = 0 \tag{4.2.10}$$
$$F(T,T,r) = 1. \tag{4.2.11}$$

By combining these, we have that

$$0 = \left(\frac{\partial A}{\partial t} - \frac{\partial B}{\partial t} r\right) F - \mu^* BF + \frac{1}{2}\sigma^2 B^2 F - rF, \tag{4.2.12}$$
$$0 = \frac{\partial A}{\partial t} - \frac{\partial B}{\partial t} r - \mu^* B + \frac{1}{2}\sigma^2 B^2 - r \tag{4.2.13}$$
$$= \frac{\partial A}{\partial t} - (1 + \frac{\partial B}{\partial t}) r - \mu^* B + \frac{1}{2}\sigma^2 B^2 \tag{4.2.14}$$

If we suppose that $\mu^*(t)$ and $\sigma^2(t)$ are affine in short-rate $r(t)$, meaning that

$$\mu^*(t) = a(t)r(t) + b(t), \tag{4.2.15}$$
$$\sigma^2(t) = c(t)r(t) + d(t), \tag{4.2.16}$$

where a,b,c and d are deterministic functions. Thus

$$0 = \frac{\partial A}{\partial t} - (1+\frac{\partial B}{\partial t})r - \mu^* B + \frac{1}{2}\sigma^2 B^2 \tag{4.2.17}$$

$$= \frac{\partial A}{\partial t} - (1+\frac{\partial B}{\partial t})r - (ar+b)B + \frac{1}{2}(cr+d)B^2 \tag{4.2.18}$$

$$= \left(\frac{\partial A}{\partial t} - bB + \frac{1}{2}dB^2\right) + \left(\frac{1}{2}cB^2 - \frac{\partial B}{\partial t} - aB - 1\right)r \tag{4.2.19}$$

and this must hold for all t,T and $r(t)$. Thus the coefficients must equal zero and we have that

$$\frac{\partial A}{\partial t} - bB + \frac{1}{2}dB^2 = 0, \tag{4.2.20}$$

$$A(T,T) = 0, \tag{4.2.21}$$

$$\frac{1}{2}cB^2 - \frac{\partial B}{\partial t} - aB - 1 = 0, \tag{4.2.22}$$

$$B(T,T) = 0. \tag{4.2.23}$$

Thus we have proved the following.

Theorem 4.2.1. *If*

$$\mu^*(t) = a(t)r + b(t), \tag{4.2.24}$$

$$\sigma^2(t) = c(t)r + d(t), \tag{4.2.25}$$

then the short-rate model has an affine term-structure model and the following equations holds:

$$\begin{cases} \frac{\partial B}{\partial t} = \frac{1}{2}cB^2 - aB - 1 \\ B(T,T) = 0 \end{cases} \tag{4.2.26}$$

$$\begin{cases} \frac{\partial A}{\partial t} - bB + \frac{1}{2}dB^2 \quad = 0 \\ A(T,T) = 0. \end{cases} \tag{4.2.27}$$

Equations of this type are Riccati Equations and they are easy to solved efficiently.

4.2.2. Vašíček–model

The material from this section is from Brigo et al. (2007, pp. 58–62).

Vašíček (1977) showed that under certain economic assumptions, the short-rate process is an Ornstein-Uhlenbeck process. This particular form was earlier suggest by

Merton (1971). Under the Vašíček–model, the short-rate process $r(t)$ is given by

$$\mathrm{d}r(t) = k(\theta - r(t))\mathrm{d}t + \sigma \mathrm{d}W(t), \tag{4.2.28}$$

where $k, \theta, \sigma > 0$ and $r(0) = r_0$ are constants. Thus

$$\mu(t, r(t)) = k(\theta - r(t)) \tag{4.2.29}$$

$$\sigma(r, t(t)) = \sigma. \tag{4.2.30}$$

Equivalent parametrization is

$$\mathrm{d}r(t) = (\theta^* - kr(t))\mathrm{d}t + \sigma \mathrm{d}W(t), \tag{4.2.31}$$

where $\theta^* = \theta k$. We could also write

$$r(t) = \theta + Y(t) \tag{4.2.32}$$

$$\mathrm{d}Y(t) = -kY(t)\mathrm{d}t + \sigma \mathrm{d}W(t). \tag{4.2.33}$$

Now $r(t) = g(t, Y(t))$, where $g(t, y) = m + y$ actually and Itô's lemma gives

$$\mathrm{d}r(t, Y(t)) = -kY(t)\mathrm{d}t + \sigma \mathrm{d}W(t) \tag{4.2.34}$$

$$= k(\theta - r(t))\mathrm{d}t + \sigma \mathrm{d}W(t). \tag{4.2.35}$$

We consider a process $X(t) = \int_0^t \mathrm{e}^{ks} \mathrm{d}W(s)$ so that $\mathrm{d}X(t) = \mathrm{e}^{kt}\mathrm{d}W(t)$. Now we define the function

$$g(x, t) = r_0 \mathrm{e}^{-kt} + \theta(1 - \mathrm{e}^{-kt}) + \sigma \mathrm{e}^{-kt} x \tag{4.2.36}$$

$$= \theta + \mathrm{e}^{-kt}(r_0 - \theta + \sigma x) \tag{4.2.37}$$

and

$$\frac{\partial g(x,t)}{\partial x} = \sigma \mathrm{e}^{-kt}, \tag{4.2.38}$$

$$\frac{\partial^2 g(x,t)}{\partial x^2} = 0, \tag{4.2.39}$$

$$\frac{\partial g(x,t)}{\partial t} = -k\mathrm{e}^{-kt}(r_0 - \theta + \sigma x) \tag{4.2.40}$$

$$= k(\theta - g(x,t)). \tag{4.2.41}$$

Since the drift term for X is zero and diffusion factor is e^{kt}, Itô's lemma for the process

$X(t)$ yields

$$
\begin{aligned}
\mathrm{d}g(X(t),t) &= \frac{\partial g(X(t),t)}{\partial t}\mathrm{d}t + \mathrm{e}^{kt}\frac{\partial g(X(t),t)}{\partial x}\mathrm{d}W && (4.2.42)\\
&= k(\theta - g(X(t),t))\mathrm{d}t + \sigma W(t). && (4.2.43)
\end{aligned}
$$

As $X(0) = 0$, we have that $g(X(0),0) = r_0$. Thus

$$
r(t) = g(X(t),t) = r_0\mathrm{e}^{-kt} + \theta(1 - \mathrm{e}^{-kt}) + \sigma\mathrm{e}^{-kt}\int_0^t \mathrm{e}^{ks}\mathrm{d}W(s). \tag{4.2.44}
$$

By Theorem A.15.1, the expected value of the integral in the equation 4.2.44 is zero and it has variance of

$$
\int_0^t \mathrm{e}^{2ks}\mathrm{d}s = \frac{1}{2k}(\mathrm{e}^{2kt} - 1). \tag{4.2.45}
$$

Hence

$$
r(t) \sim N\left(r_0\mathrm{e}^{-kt} + \theta(1 - \mathrm{e}^{-kt}), \frac{\sigma^2}{2k}(1 - \mathrm{e}^{2kt})\right). \tag{4.2.46}
$$

Since $r(t)$ is normally distributed in the Vašíček–model, there is a positive change that short-rate will be negative in a given time frame. If $t \to \infty$, then

$$
\begin{aligned}
&\mathrm{E}(r(t)) \to \theta, && (4.2.47)\\
&\mathrm{Var}(r(t)) \to \frac{\sigma^2}{2k}. && (4.2.48)
\end{aligned}
$$

We see that the parameter θ can be seen as the long-term mean and the short-rate has a tendency to move toward it. The parameter k signifies the speed of this mean-reversion while σ controls the volatility.

One of the features of the Vašíček–model is that there is a non-zero probability for negative rates. Earlier this was seen as a major drawback of the model.

Bond pricing in the Vašíček–model

If we assume that the short-rate process $r(t)$ is given by

$$
\mathrm{d}r(t) = k(\theta - r(t))\mathrm{d}t + \sigma\mathrm{d}W(t)^* \tag{4.2.49}
$$

under the risk neutral measure. Now $\mu(t) = k\theta - kr(t)$ and $\sigma(t) = \sigma$ are affine in $r(t)$. By Theorem 4.2.1, Vašíček-model has affine term-structure and

$$\begin{cases} \frac{\partial B}{\partial t} = kB - 1 \\ \frac{\partial A}{\partial t} = k\theta B - \frac{1}{2}\sigma^2 B^2 \\ A(T,T) = B(T,T) = 0, \end{cases} \tag{4.2.50}$$

Now

$$B(t,T) = \frac{1}{k}\left(1 - \mathrm{e}^{-k(T-t)}\right) \tag{4.2.51}$$

satisfies 4.2.50 and therefore we might solve $A(t,T)$ by calculation the integral

$$A(t,T) = A(t,T) - A(T,T) = -\int_t^T A(s,T)\mathrm{d}s. \tag{4.2.52}$$

We note that $B^2 = \frac{B}{k}\left(\frac{\partial B}{\partial t} + 1\right)$ and hence

$$k\theta B - \frac{1}{2}\sigma^2 B^2 = k\theta B - \frac{\sigma^2}{2k}B(1 + \frac{\partial B}{\partial t}) \tag{4.2.53}$$

$$= \frac{k^2\theta - \frac{1}{2}\sigma^2}{k}B - \frac{\sigma^2}{2k}\frac{\partial B}{\partial t}B \tag{4.2.54}$$

$$= \frac{k^2\theta - \frac{1}{2}\sigma^2}{k^2}(\frac{\partial B}{\partial t} + 1) - \frac{\sigma^2}{2k}B\frac{\partial B}{\partial t}. \tag{4.2.55}$$

Now the conditions in Equation 4.2.50 will be satisfied by

$$A(t,T) = \frac{k^2\theta - \frac{1}{2}\sigma^2}{k^2}(B(t,T) - (T-t)) - \frac{\sigma^2}{4k}B^2(t,T). \tag{4.2.56}$$

Therefore Vašíček-model has term-structure defined by

$$P(t,T) = \mathrm{e}^{A(t,T) - B(t,T)r(t)}, \tag{4.2.57}$$

where

$$B(t,T) = \frac{1}{k}\left(1 - \mathrm{e}^{-k(T-t)}\right), \tag{4.2.58}$$

$$A(t,T) = \frac{k^2\theta - \frac{1}{2}\sigma^2}{k^2}(B(t,T) - (T-t)) - \frac{\sigma^2}{4k}B^2(t,T). \tag{4.2.59}$$

Option pricing in the Vašíček–model

Since the short-rate follows a Gaussian distribution, the price of a option on a zero-coupon bond can be calculated explicitly. We shall not do that. A European call option with maturity S on a T-bond and exercise price K has a price

$$\mathrm{ZBC}(t,S,T,K) = p(t,T)N(d_1) - Kp(t,S)N(d_2), \tag{4.2.60}$$

at the time t, where

$$d_1 = \frac{\log \frac{p(t,T)}{Kp(t,S)} + \frac{V}{2}}{\sqrt{V}} \tag{4.2.61}$$

$$d_2 = \frac{\log \frac{p(t,T)}{Kp(t,S)} - \frac{V}{2}}{\sqrt{V}} \tag{4.2.62}$$

$$V = \sigma^2 \left(\frac{1 - \mathrm{e}^{-2k(T-S)}}{k} \right)^2 \frac{1 - \mathrm{e}^{-2k(S-t)}}{2k}. \tag{4.2.63}$$

A European put option with maturity S on a T-bond and exercise price K has a price

$$\mathrm{ZBP}(t,S,T,K) = Kp(t,S)N(-d_2) - p(t,T)N(-d_1) \tag{4.2.64}$$

at the time t

4.2.3. Cox-Ingersol-Ross–model (CIR)

The material from this section is mainly from Brigo et al. (2007, pp. 64–68).

Cox et al. (1985) introduced the Cox-Ingersol-Ross–model (CIR) where the short-rate process r_t is given by

$$\mathrm{d}r(t) = k(\theta - r(t))\mathrm{d}t + \sigma\sqrt{r(t)}\mathrm{d}W(t), \tag{4.2.65}$$

where r_0, k, θ and σ are positive constants. Another widely used parametrization is

$$\mathrm{d}r(t) = (\alpha - kr(t))\mathrm{d}t + \sigma\sqrt{r(t)}\mathrm{d}W(t), \tag{4.2.66}$$

where $\alpha = k\theta$. Like Vašíček–model, CIR features reversion toward the mean θ with k as the strength of the reversion. But it also has non-constant volatility as the diffusion term is $\sigma\sqrt{r(t)}$. CIR model also has affine term-structure and therefore the bond prices can be efficiently solved. Unlike Vašíček–model, CIR–model can be specified so that the short-rate will be always positive.

Bond pricing in the CIR–model

By solving the Riccati equation, which we shall not do, we get that the bond price in CIR model is

$$p(t,T) = \mathrm{e}^{A(t,T)-B(t,T)r(t)}, \tag{4.2.67}$$

where

$$A(t,T) = \frac{2k\theta}{\sigma^2}\log\left(\frac{2\beta \mathrm{e}^{\frac{(\beta+k)(T-t)}{2}}}{(\beta+k)(\mathrm{e}^{\beta(T-t)}-1)+2\beta}\right) \tag{4.2.68}$$

$$B(t,T) = \frac{2(\mathrm{e}^{\beta(T-t)}-1)}{(\beta+k)(\mathrm{e}^{\beta(T-t)}-1)+2\beta} \tag{4.2.69}$$

$$\beta = \sqrt{k^2+2\sigma^2} \tag{4.2.70}$$

Option pricing in the CIR–model

A European call option with maturity S on a T-bond and exercise price K has a price

$$\mathrm{ZBC}(t,S,T,K) = p(t,T)\chi_1^2 - Kp(t,S)\chi_2^2, \tag{4.2.71}$$

at the time t, where

$$\chi_1^2 = \chi^2\left(v_1, \frac{4k\theta}{\sigma^2}, \frac{2\beta_3^2 r(t)\mathrm{e}^{\beta(S-t)}}{\beta_2+\beta_3+B(S,T)}\right) \tag{4.2.72}$$

$$\chi_2^2 = \chi^2\left(v_2, \frac{4k\theta}{\sigma^2}, \frac{2\beta_3^2 r(t)\mathrm{e}^{\beta(S-t)}}{\beta_2+\beta_3}\right) \tag{4.2.73}$$

$$v_1 = 2(\beta_2+\beta_3+B(S,T))\frac{A(S,T)-\log(K)}{B(S,T)} \tag{4.2.74}$$

$$v_2 = 2(\beta_2+\beta_3)\frac{A(S,T)-\log(K)}{B(S,T)} \tag{4.2.75}$$

$$\beta_2 = \frac{k+\beta}{\sigma^2} \tag{4.2.76}$$

$$\beta_3 = \frac{2\beta}{\sigma^2(\mathrm{e}^{\beta(S-t)}-1)} \tag{4.2.77}$$

and $\chi^2(v,a,b)$ is the cumulative non-central chi-squared distribution with A degrees of freedom and non-centrality parameter b.

4.3. Multi-factor short-rate models

Multi-factor short-rate models have more than one state variables that drive the evolution of the short-rate. Litterman and Scheinkman (1991) demonstrated that while the majority of the yield curve movements can be explained by a single factor, it can not explain it all. Usually it is considered that at least 3 factors are needed.

One problem with one-factor affine one-factor term-structures models is that rates of different maturities are perfectly correlated. By Equation 4.2.2,

$$p(t,T) = \mathrm{e}^{A(t,T)-B(t,T)r(t)} \tag{4.3.1}$$

and the continuously compounded rate satisfies $\mathrm{e}^{R(t,T)(T-t)}p(t,T) = 1$, we see that

$$R(t,T) = \frac{B(t,T)}{T-t}r(t) - \frac{A(t,T)}{T-t}. \tag{4.3.2}$$

This implies that rates are perfectly correlated. Thus multiple factors are needed to induce realistic correlations among the rates of different maturities.

In this section we shall follow the presentation based on Nawalka, Beliaeva, and Soto (2007, pp. 425–435).

4.3.1. Simple $A(M,N)+$–models

We now define a class of models $A(M,N)+$ with $N-M$ correlated Gaussian processes and M uncorrelated square-root processes. The correlated gaussian processes are

$$\mathrm{d}Y_i(t) = -k_i Y_i(t)\mathrm{d}t + \nu_i \mathrm{d}W_i(t), \tag{4.3.3}$$

where W_i is a Wiener process and

$$\mathrm{d}W_i(t)\mathrm{d}W_j(t) = \rho_{ij}\mathrm{d}t \tag{4.3.4}$$

for all $i,j = 1,2,\dots,N-M$. Here $-1 < \rho_{ij} = \rho_{ji} < 1$ and $\rho_{ii} = 1$. The M square-root processes are

$$\mathrm{d}X_m(t) = \alpha_m(\theta_m - X_m(t))\mathrm{d}t + \sigma_m\sqrt{X_m(t)}\mathrm{d}Z_m(t) \tag{4.3.5}$$

where Z_m are independent Wiener process and

$$\mathrm{d}W_i(t)\mathrm{d}Z_m(t) = 0 \tag{4.3.6}$$

for all $i = 1,2,\dots,N-M$ and $m = 1,2,\dots,M$. The short-rate is defined by

$$r(t) = \delta + \sum_{m=1}^{M} X_m(t) + \sum_{i=1}^{N-M} Y_i(t), \tag{4.3.7}$$

where δ is a constant. Thus

$$\mathrm{d}r(t) = \left(\sum_{m=1}^{M} \alpha_m(\theta_m - X_m(t)) - \sum_{i=1}^{N-M} k_i Y_i(t) \right) \mathrm{d}t \tag{4.3.8}$$

$$+ \sum_{m=1}^{M} \sigma_m \sqrt{X_m(t)} \mathrm{d}Z_m(t) + \sum_{i=1}^{N-M} \nu_i \mathrm{d}W_i(t). \tag{4.3.9}$$

If we define

$$H(t,T) = \int_t^T \delta \mathrm{d}x = (T-t)\delta, \tag{4.3.10}$$

$$\beta_m = \sqrt{\alpha_m^2 + 2\sigma_m^2}, \tag{4.3.11}$$

$$C_i(x) = \frac{1 - \mathrm{e}^{-k_i x}}{k_i}, \tag{4.3.12}$$

$$B_m(x) = \frac{2(\mathrm{e}^{\beta_m x} - 1)}{(\beta_m + \alpha_m)(\mathrm{e}^{\beta_m x} - 1) + 2\beta_m}, \tag{4.3.13}$$

$$A(x) = \sum_{m=1}^{M} \frac{2\alpha_m \theta_m}{\sigma_m^2} \log \left(\frac{2\beta_m \mathrm{e}^{\frac{1}{2}(\beta_m + \alpha_m)x}}{(\beta_m + \alpha_m)(\mathrm{e}^{\beta_m x - 1}) + 2\beta_m} \right) \tag{4.3.14}$$

$$+ \frac{1}{2} \sum_{i=1}^{N-M} \sum_{j=1}^{N-M} \frac{\nu_i \nu_j \rho_{ij}}{k_i k_j} \left(x - C_i(x) - C_j(x) + \frac{1 - \mathrm{e}^{(k_i + k_j)x}}{k_i + k_j} \right) \tag{4.3.15}$$

for all $i = 1,2,\dots,N-M$ and $m = 1,2,\dots,M$, then the price of a zero coupon bond is given by

$$p(t,T) = \exp\left(A(\tau) - \sum_{m=1}^{M} B_m(\tau) X_m(t) - \sum_{i=1}^{N-M} C_i(\tau) Y_i(t) - H(t,T) \right), \tag{4.3.16}$$

where $\tau = T - t$. This model also has a semi-explicit formula for options on zero-coupon bonds. The method to calculate this will introduced in Section 4.5.

We note that Vašíček–model is $A(0,1)$ and CIR–model is $A(1,1)$ in this notation. We shall calibrate some of the models of class $A(M,N)+$ to market data in Chapter 6.

These $A(M,N)$–models can be made into $A(M,N)++$ models by using the dynamic extension method of the following section.

4.4. Dynamic extension to match the given term-structure

This section follows the paper by Brigo et al. (2001). This dynamic extension method allows the model to exactly fit any given the term-structure. Usually the term-structure is boot-strapped from market rates or prices and, if needed, the missing values are interpolated/extrapolated. Then the remaining parameters are are fitted to the volatility structure of some liquid option market data. The final model can be then used to price exotic options. Although we could not calibrate models to option data in emperical work in Chapter 6, we shall show how easily the +-models can be extended to ++-models using this technique.

Let $(\Omega^x, \mathbb{Q}^x, \mathscr{F}^x)$ be a probability space. We first assume the process $(x_\alpha(t))$ follows

$$\mathrm{d}x_\alpha(t) = \mu(x_\alpha(t);\alpha)\mathrm{d}t + \sigma(x_\alpha(t));\alpha)\mathrm{d}W_x(t), \tag{4.4.1}$$

$$x_\alpha(0) = x_0 \tag{4.4.2}$$

under the measure $\mathbb{Q}^x$, where α is a parameter vector. Let $\mathscr{F}_t^x$ be the member of a filtration generated by x_α up to time t. Suppose that the process $(x_\alpha(t))$ is the short-rate process under the risk-free measure $\mathbb{Q}_x$ and the price of a zero-coupon bond is

$$p^x(t,T) = \mathrm{E}_{\mathbb{Q}_x}\left(\mathrm{e}^{-\int_t^T x_\alpha(s)\ \mathrm{d}s} \mid \mathscr{F}_t^{x_\alpha}\right), \tag{4.4.3}$$

which is a function of (t, T, x_α, α). Brigo et al. (2001) calls this as a reference model. It is not guaranteed that the implied zero-curve structure by this model will match the observed market data.

Let $\varphi(t;\alpha,x_0) = \varphi(t;\alpha^*)$ be a deterministic real-valued function that it is at least integrable under any closed interval. Suppose that the short-rate follows

$$r(t) = x(t) + \varphi(t;\alpha^*) \tag{4.4.4}$$

and $(x(t))$ follows that same process under the risk-free measure $\mathbb{Q}_0$ as $(x_\alpha(t))$ does under the measure $\mathbb{Q}^x$. This model is the shifted model. This implies that

$$p(t,T) = \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^T r(s)\,\mathrm{d}s} \mid \mathscr{F}_t\right) \tag{4.4.5}$$
$$= \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^T (x(s)+\varphi(s;\alpha^*))\,\mathrm{d}s} \mid \mathscr{F}_t\right) \tag{4.4.6}$$
$$= \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^T x(s)\,\mathrm{d}s} \mid \mathscr{F}_t\right)\mathrm{e}^{-\int_t^T \varphi(s;\alpha^*)\,\mathrm{d}s} \tag{4.4.7}$$
$$= \mathrm{E}_{\mathbb{Q}^x}\left(\mathrm{e}^{-\int_t^T x_\alpha(s)\,\mathrm{d}s} \mid \mathscr{F}_t^{x_\alpha}\right)\mathrm{e}^{-\int_t^T \varphi(s;\alpha^*)\,\mathrm{d}s} \tag{4.4.8}$$
$$= p^x(t,T)\mathrm{e}^{-\int_t^T \varphi(s;\alpha^*)\,\mathrm{d}s}. \tag{4.4.9}$$

If we have a method to calculate the bond price under the reference model, then bond prices under the shifted can be calculated by a making a deterministic discounting based on the function $\varphi(t,\alpha^*)$.

In order to shorten the notations, we denote

$$I(t,T,f) = \mathrm{e}^{\int_t^T f(s)\,\mathrm{d}s}, \tag{4.4.10}$$

where f is a real valued function. If there is also a function

$$\mathrm{ZBC}^x(t,S,T,K) = \mathrm{E}_{\mathbb{Q}_x}\left(\mathrm{e}^{-\int_t^S x_\alpha(s)\,\mathrm{d}s}\left(p^x(t,T)-K\right)^+ \mid \mathscr{F}_t^{x_\alpha}\right) \tag{4.4.11}$$

for the price at time t of a call option maturing at S for a T-bond under the reference model. Now the price under the shifted model is

$$\mathrm{ZBC}(t,S,T,K) \tag{4.4.12}$$
$$= \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^S r(s)\,\mathrm{d}s}\left(p(S,T)-K\right)^+ \mid \mathscr{F}_t\right) \tag{4.4.13}$$
$$= \mathrm{e}^{-\int_t^S \varphi(s;\alpha^*)\,\mathrm{d}s}\,\mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^S x(s)\,\mathrm{d}s}\left(p(S,T)-K\right)^+ \mid \mathscr{F}_t\right) \tag{4.4.14}$$
$$= I(t,S,-\varphi)\,\mathrm{E}_{\mathbb{Q}_0}\left(I(t,S,-x)\left(p(S,T)-K\right)^+ \mid \mathscr{F}_t\right) \tag{4.4.15}$$

As

$$\mathrm{E}_{\mathbb{Q}_0}\left(I(t,S,-x)\left(p(S,T)-K\right)^+ \mid \mathscr{F}_t\right) \tag{4.4.16}$$
$$= \mathrm{E}_{\mathbb{Q}_0}\left(I(t,S,-x)\left(p^x(S,T)I(S,T,-\varphi)-K\right)^+ \mid \mathscr{F}_t\right) \tag{4.4.17}$$
$$= \mathrm{E}_{\mathbb{Q}_0}\left(I(t,S,-x)\left(p^x(S,T)-I(S,T,\varphi)K\right)^+ \mid \mathscr{F}_t\right)I(S,T,-\varphi) \tag{4.4.18}$$
$$= \mathrm{ZBC}^x(t,S,T,K\mathrm{e}^{\int_S^T \varphi(s;\alpha^*)\,\mathrm{d}s})\mathrm{e}^{-\int_S^T \varphi(s;\alpha^*)\,\mathrm{d}s}. \tag{4.4.19}$$

Here we have used again the equivalence on processes $x(t)$ and $x_{\alpha(t)}$. Hence

$$\begin{aligned} &\mathrm{ZBC}(t,S,T,K) && (4.4.20)\\ =&\mathrm{e}^{-\int_t^S \varphi(s;\alpha^*)\,\mathrm{d}s}\mathrm{ZBC}^x(t,S,T,K\mathrm{e}^{\int_S^T \varphi(s;\alpha^*)\,\mathrm{d}s})\mathrm{e}^{-\int_S^T \varphi(s;\alpha)\,\mathrm{d}s} && (4.4.21)\\ =&\mathrm{ZBC}^x(t,S,T,K\mathrm{e}^{\int_S^T \varphi(s;\alpha^*)\,\mathrm{d}s})\mathrm{e}^{-\int_t^T \varphi(s;\alpha^*)\,\mathrm{d}s} && (4.4.22)\end{aligned}$$

We see that the options prices for the shifted model can be computed easily, if the options prices can be calculated efficiently under the reference model. But not pure discounting is not enough, as we also have to shift probabilities by shifting the target strike.

Thus the following holds.

Theorem 4.4.1. *Under the earlier assumptions, the T-bond has a price*

$$p(t,T) = p^x(t,T)\mathrm{e}^{-\int_t^T \varphi(s;\alpha^*)\,\mathrm{d}s} \tag{4.4.23}$$

and a call option with maturity S on this bond has price

$$\mathrm{ZBC}(t,S,T,K) = \mathrm{ZBC}^x(t,S,T,K\mathrm{e}^{\int_S^T \varphi(s;\alpha^*)\,\mathrm{d}s})\mathrm{e}^{-\int_t^S \varphi(s;\alpha^*)\,\mathrm{d}s} \tag{4.4.24}$$

at the time t.

This model can achieve a perfect fit to the initial interest-rate term structure.

Theorem 4.4.2. *Under the earlier assumptions, the following are equivalent:*

- *The model*

$$r(t) = x(t) + \varphi(t;\alpha^*) \tag{4.4.25}$$

  has a perfect fit to the given interest-rate term structure,

- *for all $t \geq 0$*

$$\mathrm{e}^{-\int_t^T \varphi(s;\alpha^*)\mathrm{d}s} = \frac{p^M(0,T)p^x(0,t)}{p^M(0,t)p^x(0,T)}, \tag{4.4.26}$$

- *for all $t \geq 0$*

$$\varphi(t,\alpha^*) = f^M(0,t) - f^x(0,t) \tag{4.4.27}$$

where p^M are observed market prices of zero-coupon bonds, $f^M(0,t)$ is the market implied forward-rate and $f^x(0,t)$ is the forward-rate implied by the reference model.

Proof. We note that the perfect fit to market rates is equivalent to

$$p^M(0,t) = p(0,t) = \mathrm{e}^{-\int_0^t \varphi(s;\alpha^*)\,\mathrm{d}s} p^x(0,t). \tag{4.4.28}$$

That is equivalent to

$$\begin{aligned}
\mathrm{e}^{-\int_t^T \varphi(s;\alpha^*)\,\mathrm{d}s} &= \mathrm{e}^{-\int_0^T \varphi(s;\alpha^*)\,\mathrm{d}s} \mathrm{e}^{\int_0^t \varphi(s;\alpha^*)\,\mathrm{d}s} && (4.4.29)\\
&= \frac{p^M(0,T)}{p^x(0,T)} \frac{p^x(0,t)}{p^M(0,t)} && (4.4.30)
\end{aligned}$$

Equation 4.4.28 is also equivalent to

$$\log p^M(0,t) = \log p^x(0,t) - \int_0^t \varphi(s;\alpha^*)\,\mathrm{d}s \tag{4.4.31}$$

for all $t \geq 0$ and this implies that the $-f^M(0,t) = -f^x(0,t) - \varphi(t;\alpha^*)$. □

Theorem 4.4.2 guarantees that no matter how the term-structure is shaped, we can fit it exactly with a suitable function. If we want to price bonds and options on bonds in the shifted model, then the calculation of the whole function φ is unnecessary as we need only the values

$$\mathrm{e}^{-\int_t^T \varphi(s;\alpha^*)\mathrm{d}s} = \frac{p^M(0,T)p^x(0,t)}{p^M(0,t)p^x(0,T)}. \tag{4.4.32}$$

Using the market values is preferable to estimating the forward curve, as the curve fitting may cause errors.

We note that if $\varphi(t) = \varphi(t;\alpha^*)$ is differentiable, then

$$\begin{aligned}
\mathrm{d}r(t) &= \mathrm{d}x(t) + \frac{\partial}{\partial t}\varphi(t;\alpha,x_0)\mathrm{d}t && (4.4.33)\\
&= \mu(x_\alpha(t);\alpha)\mathrm{d}t + \sigma(x_\alpha(t));\alpha)\mathrm{d}W_x(t) + \frac{\partial}{\partial t}\varphi(t)\mathrm{d}t && (4.4.34)\\
&= \left(\mu(r(t)-\varphi(t);\alpha) + \frac{\partial}{\partial t}\varphi(t)\right)\mathrm{d}t + \sigma(r(t)-\varphi(t));\alpha)\mathrm{d}W_x(t). && (4.4.35)
\end{aligned}$$

We also note that if $t = 0$, then

$$p(0,T) = p^M(0,T) \tag{4.4.36}$$
$$= p^x(0,T)\mathrm{e}^{-\int_0^T \varphi(s;\alpha^*)\,\mathrm{d}s} \tag{4.4.37}$$

and

$$\mathrm{ZBC}(0,S,T,K) = \mathrm{ZBC}^x(0,S,T,K\mathrm{e}^{\int_S^T \varphi(s;\alpha^*)\,\mathrm{d}s})\mathrm{e}^{-\int_0^S \varphi(s;\alpha^*)\,\mathrm{d}s} \tag{4.4.38}$$

4.4.1. Vašíček++–model

By dynamically extending Vašíček–model, we get a model that is equivalent to a variant of Hull-White–model (J. Hull et al. 1990, Brigo et al. 2001). This section follows Brigo et al. (2007, pp. 100–102).

Suppose that we bootstrap the prices of T_i-bonds from the market for some $i = 1,2,\dots,n$. Let them be $p^M(0,T_1), p^M(0,T_2),\dots, p^M(0,T_n)$.

We assume that the evolution of $x(t)$ is given by

$$\mathrm{d}x(t) = k(\theta - x(t))\mathrm{d}t + \sigma \mathrm{d}W(t), \tag{4.4.39}$$

where $k,\theta,\sigma > 0$ and $x(0) = x_0$ are constants and $W(t)$ is a brownian motion under the risk-free measure. We make the explicit assumption that $\theta = 0$, hence

$$\mathrm{d}x(t) = -kx(t))\mathrm{d}t + \sigma \mathrm{d}W(t). \tag{4.4.40}$$

This assumption does not actually restrict the model at all. It only changes the center of the distribution, which does not matter as dynamic shift will be shifted similarly.

Now we know that

$$p^x(t,T) = \mathrm{e}^{-B(t,T)x(t)}, \tag{4.4.41}$$

where

$$B(t,T) = \frac{2(\mathrm{e}^{\beta(T-t)} - 1)}{(\beta + k)(\mathrm{e}^{\beta(T-t)} - 1) + 2\beta}, \tag{4.4.42}$$
$$\beta = \sqrt{k^2 + 2\sigma^2}. \tag{4.4.43}$$

We now define the short-rate by

$$r(t) = x(t) + \varphi(t), \tag{4.4.44}$$

where the function φ is taken as in the Theorem 4.4.2. Now

$$\mathrm{d}r(t) = \left(-kx(t) + \frac{\partial}{\partial t}\varphi(t)\right)\mathrm{d}t + \sigma \mathrm{d}W(t) \tag{4.4.45}$$

$$= \left(k(\varphi(t) - r(t)) + \frac{\partial}{\partial t}\varphi(t)\right)\mathrm{d}t + \sigma \mathrm{d}W(t). \tag{4.4.46}$$

By Theorems 4.4.1 and 4.4.2, we have that

$$p(t,T) = p^x(t,T)\mathrm{e}^{-\int_t^T \varphi(s;\alpha^*)\ \mathrm{d}s} \tag{4.4.47}$$

$$= p^x(t,T)\frac{p^M(0,T)p^x(0,t)}{p^M(0,tp^x(0,T)} \tag{4.4.48}$$

$$= \frac{p^M(0,T)}{p^M(0,t)}\mathrm{e}^{-B(t,T)x(t)-B(0,t)x(0)+B(0,T)x(0)} \tag{4.4.49}$$

if $t = T_i$ and $T = T_j$ for some $i < j$. If $t = 0$, then $p(0,T) = p^M(0,T)$, as was expected.

A European call option with maturity S on a T-bond and exercise price K has a price

$$\mathrm{ZBC}^x(t,S,T,K) = p^x(t,T)N(d_1) - Kp^x(t,S)N(d_2), \tag{4.4.50}$$

at the time t under the dynamics of $x(t)$, where

$$d_1 = \frac{\log\frac{p^x(t,T)}{Kp^x(t,S)} + \frac{V}{2}}{\sqrt{V}} \tag{4.4.51}$$

$$d_2 = \frac{\log\frac{p^x(t,T)}{Kp^x(t,S)} - \frac{V}{2}}{\sqrt{V}} \tag{4.4.52}$$

$$V = \sigma^2\left(\frac{1-\mathrm{e}^{-2k(T-S)}}{k}\right)^2\frac{1-\mathrm{e}^{-2k(S-t)}}{2k}. \tag{4.4.53}$$

Now, by Theorems 4.4.1 and 4.4.2, we have that

$$\mathrm{ZBC}(t,S,T,K) = \mathrm{ZBC}^x(t,S,T,K\mathrm{e}^{\int_S^T \varphi(s;\alpha^*)\ \mathrm{d}s})\mathrm{e}^{-\int_t^S \varphi(s;\alpha^*)\ \mathrm{d}s} \tag{4.4.54}$$

$$= \mathrm{ZBC}^x(t,S,T,K/I(S,T))I(t,S), \tag{4.4.55}$$

where

$$I(t,S) = \frac{p^M(0,S)p^x(0,t)}{p^M(0,t)p^x(0,S)} \tag{4.4.56}$$

$$= \frac{p^M(0,S)}{p^M(0,t)} \mathrm{e}^{-B(0,t)x(0)+B(0,S)x(0)} \tag{4.4.57}$$

and

$$I(S,T) = \frac{p^M(0,T)p^x(0,S)}{p^M(0,S)p^x(0,T)} \tag{4.4.58}$$

$$= \frac{p^M(0,T)}{p^M(0,S)} \mathrm{e}^{-B(0,S)x(0)+B(0,T)x(0)}. \tag{4.4.59}$$

4.4.2. CIR++–model

Extended CIR–model, or CIR++–model, can be constructed by combining the formulas in section 4.2.3 with the results from section 4.4 as was done in section 4.4.1. As this is trivial, we shall writing the formulas again.

4.4.3. G2++–model

$A(2,2)++$–model is achieved by extending $A(2,2)$–model as above. Thus model is equivalent to model by J. C. Hull and White (1994) and it is called as $G2++$–model by Brigo et al. (2007).

Brigo et al. (2007, pp. 153–156, 172–173) contains the proof for the following.

Theorem 4.4.3. *A European call option with maturity S on a T-bond and exercise price K has a price*

$$\mathrm{ZBC}^x(t,S,T,K) = p(t,T)\Phi\left(K^* + \frac{1}{2}S(t,S,T)\right) \tag{4.4.60}$$

$$-\,p(t,S)K\Phi\left(K^* - \frac{1}{2}S(t,S,T)\right) \tag{4.4.61}$$

where

$$K^* = \frac{\log \frac{p(t,T)}{Kp(t,S)}}{S(t,S,T)} \tag{4.4.62}$$

$$S(t,S,T)^2 = \sum_{i=1}^{2} \frac{\nu_i^2}{2k_i^3}\left(1-\mathrm{e}^{-k_i(T-S)}\right)^2\left(1-\mathrm{e}^{-2k_i(S-t)}\right) \tag{4.4.63}$$

$$+\,2\rho\frac{\nu_1\nu_2}{k_1k_2(k_1+k_2)}\left(1-\mathrm{e}^{-k_1(T-S)}\right)\left(1-\mathrm{e}^{-k_2(T-S)}\right)\left(1-\mathrm{e}^{-(k_1+k_2)(S-t)}\right) \tag{4.4.64}$$

4.5. Option valuation using Fourier inversion method

Now we show how to value options using the Fourier inversion method introduced by Heston (1993). Our approach is based on Nawalka et al. (2007, pp. 222–233).

We make the following assumptions. The interest rate model is affine $A_M(N)$ model. Thus

$$r(t) = \delta(t) + \sum_{m=1}^{M} X_m(t) + \sum_{i=1}^{N-M} Y_i(t). \tag{4.5.1}$$

The correlated gaussian process are

$$\mathrm{d}Y_i(t) = -k_i Y_i(t)\mathrm{d}t + \nu_i \mathrm{d}W_i(t), \tag{4.5.2}$$

where W_i is a Wiener process and

$$\mathrm{d}W_i(t)\mathrm{d}W_j(t) = \rho_{ij}\mathrm{d}t \tag{4.5.3}$$

for all $i, j = 1, 2, \ldots, N-M$. Here $-1 < \rho_{ij} = \rho_{ji} < 1$ and $\rho_{ii} = 1$. The M square-root processes are

$$\mathrm{d}X_m(t) = \alpha_m(\theta_m - X_m(t))\mathrm{d}t + \sigma_m\sqrt{X_m(t)}\mathrm{d}Z_m(t) \tag{4.5.4}$$

where Z_m are independent Wiener process and

$$\mathrm{d}W_i(t)\mathrm{d}Z_m(t) = 0 \tag{4.5.5}$$

for all $i = 1, 2, \ldots, N-M$ and $m = 1, 2, \ldots, M$.

We explicitly assume that $\delta(t) = 0$ for all $t \geq 0$ to keep notation simple and we do the dynamic extension of Section 4.4 to add the $\delta(t)$ afterwards. Now zero-coupon bond prices are given by

$$p(t,T) = \mathrm{e}^{A(\tau) - B^\top(\tau)X(t) - C^\top(\tau)Y(t)}, \tag{4.5.6}$$

where $\tau = T - t$. This is a simplification from Nawalka et al. (2007, p. 433–435).

The price of a call option expiring on S written on T-bond with strike price K is

$$
\begin{aligned}
c(t) &= \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^S r(s)\mathrm{d}s}\left(p(S,T)\right)-K\right)_+ \mid \mathscr{F}_t\right) && (4.5.7)\\
&= \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{-\int_t^S r(s)\mathrm{d}s}\left(p(S,T)\right)-K\right) 1_{p(S,T)-K\geq 0} \mid \mathscr{F}_t\right) && (4.5.8)\\
&= p(t,T)\Pi_1 - Kp(t,S)\Pi_2, && (4.5.9)
\end{aligned}
$$

where

$$
\begin{aligned}
\Pi_1 &= \mathrm{E}_{\mathbb{Q}_0}\left(\frac{\mathrm{e}^{-\int_t^S r(s)\mathrm{d}s} p(S,T) 1_{p(S,T)\geq K}}{p(t,T)} \mid \mathscr{F}_t\right) && (4.5.10)\\
\Pi_2 &= \mathrm{E}_{\mathbb{Q}_0}\left(\frac{\mathrm{e}^{-\int_t^S r(s)\mathrm{d}s} 1_{p(S,T)\geq K}}{p(t,T)} \mid \mathscr{F}_t\right). && (4.5.11)
\end{aligned}
$$

Here all the expectations are taken under the risk-free measure. We now write Π_1 under a different measure. As $p(t,T) \geq K$ is equivalent to $\ln p(t,T) \geq \ln K$, we change the variable by $y = \ln p(t,T)$ and get

$$
\Pi_1 = \int_{\ln K}^{\infty} \left(\frac{\mathrm{e}^{-\int_t^S r(s)\mathrm{d}s} p(S,T)}{p(t,T)} f(y)\right) \mathrm{d}y. \tag{4.5.12}
$$

We notice that

$$
\begin{aligned}
\xi_1(t) &= \frac{p(S,T)B(t)}{p(t,T)B(S)} && (4.5.13)\\
&= \frac{p(S,T)}{p(t,T)} \mathrm{e}^{-\int_t^S r(s)\mathrm{d}s}. && (4.5.14)
\end{aligned}
$$

is the Radon-Nikodým derivative of T-forward measure with respect to risk-free measure. Thus

$$
\begin{aligned}
\Pi_1 &= \int_{\ln K}^{\infty} \xi_1(t) f(y)\, \mathrm{d}y && (4.5.15)\\
&= \int_{\ln K}^{\infty} f_1(y)\, \mathrm{d}y && (4.5.16)
\end{aligned}
$$

where f_1 is the probability density function under the T-forward measure. Let g_1 be the characteristic function of T-forward measure, hence

$$g_1(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega y} f_1(y)\ \mathrm{d}y \tag{4.5.17}$$

$$= \int_{-\infty}^{\infty} \left(\mathrm{e}^{i\omega y}\xi_1(t)f(y)\right)\ \mathrm{d}y \tag{4.5.18}$$

$$= \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{i\omega \ln p(S,T)}\xi_1(t) \mid \mathscr{F}_t\right) \text{ and} \tag{4.5.19}$$

$$g_1(\omega)p(S,T) = \mathrm{E}_{\mathbb{Q}_0}\left(\mathrm{e}^{(1+i\omega)\ln p(S,T)}\mathrm{e}^{-\int_t^S r(s)\mathrm{d}s} \mid \mathscr{F}_t\right) \tag{4.5.20}$$

as, by Equation 4.5.6,

$$y = \ln p(S,T) = A(\tau) - B^{\top}(\tau)X(t) - C^{\top}(\tau)Y(t). \tag{4.5.21}$$

Feynman-Kac theorem this expected value may be presented as a N-dimensional stochastic partial differential equation. Nawalka et al. (2007) shows that a solution is

$$\exp\left(A_1^*(s) - \sum_{m=1}^{M} B_{1m}^*(s)X_m(t) - \sum_{i=1}^{N-M} C_{1m}^*(s)Y_i(t)\right), \tag{4.5.22}$$

where

$$A_1^*(0) = a_1 = A(U)(1+i\omega) \tag{4.5.23}$$

$$B_{1m}^*(0) = b_{1m} = B_m(U)(1+i\omega) \tag{4.5.24}$$

$$C_{1i}^*(0) = c_{1i} = C_i(U)(1+i\omega) \tag{4.5.25}$$

for $i = 1,2,\ldots,N-M$ and $m = 1,2,\ldots,M$. Here

$$A_1^*(z) = a_1 + \frac{1}{2}\sum_{i=1}^{N-M}\sum_{j=1}^{N-M}\frac{\nu_i\nu_j\rho_{ij}}{k_ik_j}(z - q_iC_i(z) - q_jC_j(z) \tag{4.5.26}$$

$$+ q_iq_j\frac{1-\exp^{-(k_i+k_j)z}}{k_i+k_j}) \tag{4.5.27}$$

$$-2\sum_{m=1}^{M}\frac{\alpha_m\theta_m}{\sigma_m^2}\left(\beta_{3m}z + \log(\frac{1-\beta_{4m}\exp^{\beta_{1m}z}}{1-\beta_{4m}})\right) \tag{4.5.28}$$

$$B_{1m}^*(z) = \frac{2}{\sigma_m^2}\left(\frac{\beta_{2m}\beta_{4m}\exp^{\beta_{1m}z} - \beta_{3m}}{\beta_{4m}\exp^{\beta_{1m}z} - 1}\right) \tag{4.5.29}$$

$$C_{1i}^*(z) = \frac{1-q_i\exp^{-k_iz}}{k_i}, \tag{4.5.30}$$

where

$$q_i = 1 - kc_{1j} \tag{4.5.31}$$

$$\beta_{1m} = \sqrt{\alpha_m^2 + 2\sigma_m^2} \tag{4.5.32}$$

$$\beta_{2m} = \frac{-\alpha_m + \beta_{1m}}{2} \tag{4.5.33}$$

$$\beta_{3m} = \frac{-\alpha_m - \beta_{1m}}{2} \tag{4.5.34}$$

$$\beta_{4m} = \frac{-\alpha_m - \beta_{1m} - b_{1m}\sigma_m^2}{-\alpha_m + \beta_{1m} - b_{1m}\sigma_m^2} \tag{4.5.35}$$

for $i = 1,2,\ldots,N-M$ and $m = 1,2,\ldots,M$. It should be noted that a_1, b_{1m} and c_{1i} are actually functions of ω and therefore A_1^*, B_{1m}^* and C_{1i}^* also depend on ω.

Thus the characteristic function is

$$g_1(\omega) = \frac{\exp\left(A_1^*(s) - \sum_{m=1}^{M} B_{1m}^*(s)X_m(t) - \sum_{i=1}^{N-M} C_{1m}^*(s)Y_i(t)\right)}{p(t,T)} \tag{4.5.36}$$

which allows us to calculate the values of characteristic functions. Now

$$\Pi_1 = \int_{\ln K}^{\infty} f_1(y)\,\mathrm{d}y \tag{4.5.37}$$

$$= \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \Re\left(\frac{\exp^{-i\omega\log K} g_1(\omega)}{i\omega}\right)\mathrm{d}\omega \tag{4.5.38}$$

which can be calculated numerically. Nawalka et al. (2007) note that this computation only requires that the model has analytical bond pricing formulas, so it can be utilized in variety of models.

We can solve Π_2 similarly. Instead of Equation 4.5.12, we have

$$\Pi_2 = \int_{\ln K}^{\infty} \left(\frac{\mathrm{e}^{-\int_t^S r(s)\mathrm{d}s} p(S,T)}{p(t,T)} f(y)\right)\mathrm{d}y. \tag{4.5.39}$$

and similar reasoning shows that

$$\Pi_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \Re\left(\frac{\exp^{-i\omega\log K} g_2(\omega)}{i\omega}\right)\mathrm{d}\omega. \tag{4.5.40}$$

Now

$$g_2(\omega) = \frac{\exp\left(A_2^*(s) - \sum_{m=1}^{M} B_{2m}^*(s)X_m(t) - \sum_{i=1}^{N-M} C_{1m}^*(s)Y_i(t)\right)}{p(t,S)} \tag{4.5.41}$$

where

$$A_2^*(0) = a_2 = A(U)(i\omega) \tag{4.5.42}$$
$$B_{2m}^*(0) = b_{2m} = B_m(U)(i\omega) \tag{4.5.43}$$
$$C_{2i}^*(0) = c_{2i} = C_i(U)(i\omega) \tag{4.5.44}$$

for $i = 1,2,\ldots,N-M$ and $m = 1,2,\ldots,M$. Similarly as before

$$A_2^*(z) = a_2 + \frac{1}{2}\sum_{i=1}^{N-M}\sum_{j=1}^{N-M}\frac{\nu_i\nu_j\rho_{ij}}{k_ik_j}(z - q_iC_i(z) - q_jC_j(z) \tag{4.5.45}$$
$$+ q_iq_j\frac{1-\exp^{-(k_i+k_j)z}}{k_i+k_j}) \tag{4.5.46}$$
$$-2\sum_{m=1}^{M}\frac{\alpha_m\theta_m}{\sigma_m^2}\left(\beta_{3m}z + \log(\frac{1-\beta_{4m}\exp^{\beta_{1m}z}}{1-\beta_{4m}})\right) \tag{4.5.47}$$
$$B_{2m}^*(z) = \frac{2}{\sigma_m^2}\left(\frac{\beta_{2m}\beta_{4m}\exp^{\beta_{1m}z} - \beta_{3m}}{\beta_{4m}\exp^{\beta_{1m}z} - 1}\right) \tag{4.5.48}$$
$$C_{2i}^*(z) = \frac{1-q_i\exp^{-k_iz}}{k_i}, \tag{4.5.49}$$

where

$$q_i = 1 - kc_{2j} \tag{4.5.50}$$
$$\beta_{1m} = \sqrt{\alpha_m^2 + 2\sigma_m^2} \tag{4.5.51}$$
$$\beta_{2m} = \frac{-\alpha_m + \beta_{2m}}{2} \tag{4.5.52}$$
$$\beta_{3m} = \frac{-\alpha_m - \beta_{2m}}{2} \tag{4.5.53}$$
$$\beta_{4m} = \frac{-\alpha_m - \beta_{2m} - b_{2m}\sigma_m^2}{-\alpha_m + \beta_{2m} - b_{2m}\sigma_m^2} \tag{4.5.54}$$

for $i = 1,2,\ldots,N-M$ and $m = 1,2,\ldots,M$. Again, it should be noted that a_2, b_{2m} and c_{2i} are actually functions of ω and therefore A_2^*, B_{2m}^* and C_{2i}^* also depend on ω.

Since the computations above involves numerical integration, it is computationally costly. However Carr and Madan (1999) showed that Fast Fourier Transform (FFT) can utilized in the computation, which can reduce the complexity significantly. Sadly, this was not attempted in this thesis.

As we explicitly assumed that $\delta = 0$, we can add the shift to the model by the extension method introduced in Section 4.4.

5. INTENSITY MODELS

A rigorous treatment of intensity modeling can be found in Bielecki and Rutkowski (2002). Brigo et al. (2007) contains a gentler introduction.

5.1. Foundations of intensity models

5.1.1. Introduction

This section is based on Brigo et al. (2007, pp. 759–764).

Let $(\Omega, \mathscr{F}, \mathbb{Q})$ be a fixed probability space and ζ a random time, meaning that $\zeta : \Omega \to \mathbb{R}_+ \cup \{0\}$ is a measurable random variable. In our setting $\zeta(\omega)$ will be the time of the default. The corresponding indicator variable is $H_t = 1_{\{\zeta \leq t\}}$. Hence, $H_t = 0$ before the default and $H_t = 1$ at the default and after it.

If we assume that ζ is exponentially distributed with parameter $\gamma > 0$, then

$$\mathbb{Q}(\zeta > t) = \mathrm{e}^{-\gamma t}. \tag{5.1.1}$$

We note that the random variable $\zeta\gamma$ is exponentially distributed with parameter 1, since

$$\mathbb{Q}(\zeta\gamma > t) = \mathbb{Q}(\zeta > \frac{t}{\gamma}) = \mathrm{e}^{-t}. \tag{5.1.2}$$

Now

$$\begin{aligned}
\mathbb{Q}(\zeta > t + \mathrm{d}t \mid \zeta > t) &= \frac{\mathbb{Q}(\zeta > t + \mathrm{d}t)}{\mathbb{Q}(\zeta > t)} && (5.1.3)\\
&= \mathrm{e}^{-\gamma(t+\mathrm{d}t)}\mathrm{e}^{\gamma t} && (5.1.4)\\
&= \mathrm{e}^{-\gamma \mathrm{d}t} && (5.1.5)\\
&= \mathbb{Q}(\zeta > \mathrm{d}t), && (5.1.6)
\end{aligned}$$

which implies that the distribution has no memory. While memorylessness is often a desirable property, time homogeneous is not a desirable property in credit risk modeling.

However, we can easily see the theoretical connection between interest rate and default intensity using a simplified model, where the risk-free rate r and the parameter γ are positive constants. Now the zero-coupon bond with maturity T and no recovery value at the default has the expected discounted value of

$$\mathbb{Q}(\zeta > t)\mathrm{e}^{-rT} + \mathbb{Q}(\zeta \leq t) \cdot 0 = \mathrm{e}^{-(r+\gamma)T}. \tag{5.1.7}$$

Thus the parameter $\gamma > 0$ can be seen as a credit spread over the risk-free rate.

We can generalize the equation 5.1.1 and introduce time-dependency by setting

$$\mathbb{Q}(\zeta > t) = \mathrm{e}^{-\Gamma(t)}, \tag{5.1.8}$$

where $\Gamma(t)$ is the cumulative hazard function. Intuitively we assume that Γ is a strictly increasing function. Now if $r(t)$ is a deterministic short-rate, then a defaultable zero-coupon bond with maturity T and no recovery value has expected discounted value of

$$\mathrm{e}^{-\left(\int_0^T r(s)\mathrm{d}s + \Gamma(T)\right)}. \tag{5.1.9}$$

As in equation 5.1.2, if $\xi = \Gamma(\zeta)$, then

$$\mathbb{Q}(\Gamma(\zeta) > t) = \mathbb{Q}(\zeta > \Gamma^{-1}(t)) = \mathrm{e}^{-t} \tag{5.1.10}$$

meaning that ξ is exponentially distributed with parameter 1 and $\zeta \sim \Gamma^{-1}(\xi)$. Now we may simulate the default time by drawing a realization of ξ and taking $\zeta = \Gamma^{-1}(\xi)$.

Next we assume that

$$\Gamma(t) = \int_0^t \gamma(s)\mathrm{d}s, \tag{5.1.11}$$

where $\gamma > 0$ almost everywhere. Now the equation 5.1.9 can be written as

$$\mathrm{e}^{-\int_0^T (r(s)+\gamma(s))\mathrm{d}s} \tag{5.1.12}$$

and again $\gamma(t)$ can be viewed as a credit spread over the risk-free rate. Now

$$\begin{aligned}
\mathbb{Q}(t \le \zeta < t + \mathrm{d}t) &= \mathrm{e}^{-\Gamma(t)} - \mathrm{e}^{-\Gamma(t+\mathrm{d}t)} && (5.1.13)\\
&= \mathrm{e}^{-\Gamma(t)} \left(1 - \mathrm{e}^{-\int_t^{t+\mathrm{d}t} \gamma(s)\mathrm{d}s}\right) && (5.1.14)\\
&\approx \mathrm{e}^{-\Gamma(t)} \int_t^{t+\mathrm{d}t} \gamma(s)\mathrm{d}s && (5.1.15)\\
&\approx \mathrm{e}^{-\Gamma(t)} \lambda(t)\mathrm{d}t && (5.1.16)\\
&= \mathrm{e}^{-\int_0^t \gamma(s)\mathrm{d}s} \lambda(t)\mathrm{d}t, && (5.1.17)
\end{aligned}$$

where the first approximation uses $\mathrm{e}^x \approx 1 + x$ given $x \approx 0$ and the second is based on the definition of the integral and the assumption that λ.

Similarly the conditional probability has the following approximation

$$\begin{aligned}
\mathbb{Q}(\zeta \le t + \mathrm{d}t | \zeta > t) &= \frac{\mathbb{Q}(t < \zeta \le t + \mathrm{d}t)}{\mathbb{Q}(\zeta > t)} && (5.1.18)\\
&= \frac{\mathrm{e}^{-\Gamma(t)} - \mathrm{e}^{-\Gamma(t+\mathrm{d}t)}}{\mathrm{e}^{-\Gamma(t)}} && (5.1.19)\\
&= 1 - \mathrm{e}^{-\int_t^{t+\mathrm{d}t} \gamma(s)\mathrm{d}s} && (5.1.20)\\
&\approx \gamma(t)\mathrm{d}t && (5.1.21)
\end{aligned}$$

Suppose that F is the cumulative distribution function of ζ, so

$$F(t) = \mathbb{Q}(\zeta \le t), \tag{5.1.22}$$

and the function F is absolutely continuous. This means that the derivative of F exists and $F' = f$ almost everywhere, where f is the density funtion of ζ. We denote $\bar{F}(t) = 1 - F(t) = \mathbb{Q}(\zeta > t)$ and make the following assumptions

(A) $\mathbb{Q}(\zeta = t) = F(0) = 0$ for all $t \ge 0$.

(B) $F(t) < 1$ for all $0 \le t < \infty$.

Now we may use Bayes rule to see that

$$
\begin{aligned}
\frac{\mathbb{Q}(\zeta \leq t+\mathrm{d}t|\zeta > t)}{\mathrm{d}t} &= \frac{\mathbb{Q}(t < \zeta \leq t+\mathrm{d}t)}{\mathbb{Q}(\zeta > t)\mathrm{d}t} &(5.1.23)\\
&= \frac{F(t+\mathrm{d}t)-F(t)}{\mathbb{Q}(\zeta > t)\mathrm{d}t} &(5.1.24)\\
&\longrightarrow \frac{f(t)}{\mathbb{Q}(\zeta > t)} &(5.1.25)\\
&= \frac{f(t)}{\bar{F}(t)} &(5.1.26)\\
&= -\frac{\mathrm{d}}{\mathrm{d}t}\log(\bar{F}(t)) &(5.1.27)\\
&= \frac{\mathrm{d}}{\mathrm{d}t}\Gamma(t) &(5.1.28)\\
&= \gamma(t) &(5.1.29)
\end{aligned}
$$

as $\mathrm{d}t \to 0^+$.

We note that

$$
\begin{aligned}
\mathrm{d}\mathbb{Q}(\zeta > t) &= -\lambda(t)\mathrm{e}^{-\int_0^t \gamma(s)\mathrm{d}s} &(5.1.30)\\
&= -\lambda(t)\mathrm{d}\mathbb{Q}(\zeta > t) &(5.1.31)
\end{aligned}
$$

holds.

In summary,

$$
\begin{aligned}
\gamma(t) &= \Gamma'(t) &(5.1.32)\\
\Gamma(t) &= \int_0^t \gamma(s)\mathrm{d}s &(5.1.33)\\
\mathbb{Q}(\zeta > t) &= \mathrm{e}^{-\Gamma(t)} = \mathrm{e}^{-\int_0^t \gamma(s)\mathrm{d}s} &(5.1.34)\\
\gamma(t)\mathrm{d}t &\approx \mathbb{Q}(\zeta \leq t+\mathrm{d}t|\zeta > t). &(5.1.35)
\end{aligned}
$$

In general setting, the function $\gamma(t)$ is the hazard function of ζ. The hazard function can be seen as the instantaneous probability of default happening just after the time t given the survival up to time t. In the context of credit risk, we shall call the function γ as the intensity function.

If $\lambda(t) = \lambda > 0$ is a deterministic constant, then F is the cumulative distribution function of exponential distribution with parameter λ and therefore $\mathrm{E}(\zeta) = 1/\lambda$ and $\mathrm{Var}(\zeta) = 1/\lambda^2$. Also

$$
\mathrm{E}(1_{\{\zeta > t\}}) = \mathbb{Q}(\zeta > t) = \mathrm{e}^{-\int_0^t \lambda_s \mathrm{d}s} = \mathrm{e}^{-\lambda t}. \tag{5.1.36}
$$

Thus ζ is signaled by the first jump of time-homogenous Poisson distribution with parameter λ. Similarly, if $\lambda(t) > 0$ is deterministic function, then ζ is the first jump of non-homogenous Poisson distribution with rate function $\lambda(t)$. If $\lambda(t)$ is a stochastic process, then ζ will follow a Cox process.

5.1.2. The credit triangle

The derivation of the credit triangle follows O'Kane (2011, pp. 54–55) although the note after the credit triangle might be original. The credit triangle shows how the default intensity can be seen approximating the spread of a credit default swap.

In this section we consider a simplified CDS contract in the intensity framework. We assume the following

1. $\lambda(t) = \lambda > 0$ is a deterministic constant.
2. The timing of default ζ is independent from interest rates under the measure $\mathbb{Q}$.
3. The recovery rate $0 \leq \mathrm{R_{EC}} \leq 1$ is a deterministic constant and it is paid at the moment of the default.
4. CDS with no upfront costs pays premium continuously at rate s until the default ζ or the termination date T.

The last assumption means that in the interval $[t, t + \mathrm{d}t]$ the paid premium is $s\mathrm{d}t$ and if $\mathrm{d}t$ is tiny, then the present value of this is $p(0,t)s\mathrm{d}t$. The value of the premium leg is then

$$\mathrm{E}\left(\int_0^T D(0,t)s1_{\{\zeta>t\}}\mathrm{d}t\right) = s\int_0^T \mathrm{E}\left(D(0,t)1_{\{\zeta>t\}}\right)\mathrm{d}t \tag{5.1.37}$$

$$= s\int_0^T p(0,t)\mathbb{Q}(\zeta > t)\mathrm{d}t. \tag{5.1.38}$$

For the valuation of the protection leg, we calculate

$$\mathrm{E}\left(D(0,\zeta)(1-\mathrm{R_{EC}})1_{\{\zeta\leq T\}}\right) = (1-\mathrm{R_{EC}})\mathrm{E}\left(\int_0^T D(0,t)1_{\{t\leq\zeta<\zeta+\mathrm{d}t\}}\right) \tag{5.1.39}$$

$$= \mathrm{L_{GD}}\int_0^T \mathrm{E}\left(D(0,t)1_{\{t\leq\zeta<\zeta+\mathrm{d}t\}}\right) \tag{5.1.40}$$

$$= \mathrm{L_{GD}}\int_0^T p(0,t)\mathbb{Q}(t\leq\zeta<t+\mathrm{d}t) \tag{5.1.41}$$

$$= -\mathrm{L_{GD}}\int_0^T p(0,t)\mathrm{d}\mathbb{Q}(\zeta>t), \tag{5.1.42}$$

where in the last step we used the derivative of the identity $\mathbb{Q}(\zeta \le t) = 1 - \mathbb{Q}(\zeta > t)$.

Now

$$\mathrm{E}\left(D(0,\zeta)\mathrm{L_{GD}}1_{\{\zeta \le T\}}\right) = \mathrm{L_{GD}}\lambda \int_0^T p(0,t)\mathbb{Q}(\zeta > t)\mathrm{d}t \tag{5.1.43}$$

and since this must be equal to value in equation (5.1.37), we get the following identity

$$s = \lambda \mathrm{L_{GD}}. \tag{5.1.44}$$

This is the credit triangle. It is also quick and easy to understand, but only one of the three variables are actually directly observable from market data. If $\mathrm{R_{EC}} = 0$, then $s = \lambda$ and the default intensity is the coupon rate intensity of the CDS.

It should be noted that the model has pathological behavior if $\mathrm{R_{EC}} \approx 1$. If $\mathrm{R_{EC}} = 1$, then a defaultable zero coupon bond is more valuable than otherwise identical risk-free bond since the defaultable bond might pay the principal earlier[1]. Since the recovery value is rarely near the notional value, this is not a serious problem.

Nowadays credit default swaps are traded with standardized coupon rates and upfront payments. However, if the CDS has a upfront value of U, then

$$U = s\int_0^T p(0,t)\mathbb{Q}(\zeta > t)\mathrm{d}t - \mathrm{L_{GD}}\lambda \int_0^T p(0,t)\mathbb{Q}(\zeta > t)\mathrm{d}t \tag{5.1.45}$$

$$= (s - \lambda \mathrm{L_{GD}})Q(r,\lambda), \tag{5.1.46}$$

where

$$Q(r,\lambda) = \int_0^T p(0,t)\mathbb{Q}(\zeta > t)\mathrm{d}t. \tag{5.1.47}$$

If we assume that the short-rate is roughly a constant r for all times $0 < t < T$, then

$$Q(r,\lambda) \approx \int_0^T \mathrm{e}^{-(r+\lambda)t}\mathrm{d}t \tag{5.1.48}$$

$$= \frac{1-\mathrm{e}^{-(r+\lambda)T}}{r+\lambda} \tag{5.1.49}$$

and thus

$$U = (s - \lambda \mathrm{L_{GD}})\frac{1-\mathrm{e}^{-(r+\lambda)T}}{r+\lambda}. \tag{5.1.50}$$

[1]One reasonable restriction that will preclude this is $\mathrm{R_{EC}} < p(0,T)$

5.2. Pricing

The pricing argumentation follows closely Brigo et al. (2007, pp. 790–792).

In this section we assume that the σ-algebra $(\mathscr{F}_t)$ presents partial market information without default and

$$\mathscr{H}_t = \sigma(1_{\{\zeta \leq s\}} | s \leq t) = \sigma(H(s) | s \leq t) \tag{5.2.1}$$

is the knowledge of the default up to time t. By

$$\mathscr{G}_t = \mathscr{F}_t \vee \mathscr{H}_t \tag{5.2.2}$$

we denote the smallest σ-algebra containing $\mathscr{F}_t$ and $\mathscr{H}_t$. We assume that conditions (DS1) and (DS2) of Section A.14 are satisfied by the process λ.

Theorem A.14.1 is an important tool in our arsenal, so we restate it here. Under very reasonable assumptions, we have that

$$\mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta > T\}} X | \mathscr{G}_t\right) = 1_{\{\zeta > t\}} \mathrm{e}^{\int_0^t \lambda(s)\mathrm{d}s}\left(1_{\{\zeta > T\}} X | \mathscr{F}_t\right) \tag{5.2.3}$$

for random variables X and $T \geq t$.

Defaultable zero coupon bond with no recovery

A defaultable T-bond with no recovery has pay-off $H_T = 1_{\{\zeta > T\}}$. Now by Lemma A.14.1,

$$\begin{aligned}
d_0(t,T) =& \mathrm{E}_{\mathbb{Q}}\left(\exp^{-\int_t^T r(s)\mathrm{d}s} H_T \mid \mathscr{F}_t \vee \mathscr{H}_t\right) && (5.2.4)\\
=& 1_{\{\zeta > t\}} \exp^{\int_0^t \lambda(s)\mathrm{d}s} \mathrm{E}_{\mathbb{Q}}\left(\exp^{-\int_t^T r(s)\mathrm{d}s} H_T \mid \mathscr{F}_t\right) && (5.2.5)\\
=& 1_{\{\zeta > t\}} \exp^{\int_0^t \lambda(s)\mathrm{d}s} \mathrm{E}_{\mathbb{Q}}\left(\exp^{-\int_t^T r(s)\mathrm{d}s} \mathrm{E}_{\mathbb{Q}}\left(H_T | \mathscr{F}_T\right) \mid \mathscr{F}_t\right) && (5.2.6)\\
=& 1_{\{\zeta > t\}} \exp^{\int_0^t \lambda(s)\mathrm{d}s} \mathrm{E}_{\mathbb{Q}}\left(\exp^{-\int_t^T r(s)\mathrm{d}s} \exp^{-\int_0^T \lambda(s)\mathrm{d}s} \mid \mathscr{F}_t\right) && (5.2.7)\\
=& 1_{\{\zeta > t\}} \mathrm{E}_{\mathbb{Q}}\left(\exp^{-\int_t^T (r(s)+\lambda(s))\mathrm{d}s} \mid \mathscr{F}_t\right) && (5.2.8)
\end{aligned}$$

If $\lambda(s) \geq 0$ almost surely, then we may see

$$r(s) + \lambda(s) \geq r(s) \tag{5.2.9}$$

is the defaultable short-rate. Thus we may reuse all the machinery from the short-rate models.

Defaultable zero coupon bond with partial recovery $0 < \mathrm{R_{EC}} < 1$ at the maturity

A defaultable zero coupon bond with maturity T and partial recovery at the maturity has pay-off

$$
\begin{aligned}
1_{\{\zeta>T\}} + \mathrm{R_{EC}} 1_{\{\zeta\le T\}} &= (1-\mathrm{R_{EC}})1_{\{\zeta>T\}} + \mathrm{R_{EC}} && (5.2.10)\\
&= 1_{\{\zeta>T\}}\mathrm{L_{GD}} + \mathrm{R_{EC}} && (5.2.11)
\end{aligned}
$$

at the maturity. Thus the price of it at the time t is

$$
d_M(t,T) = d_0(t,T)\mathrm{L_{GD}} + p(t,T)\mathrm{R_{EC}}, \tag{5.2.12}
$$

where d_0 is the price of defaultable zero coupon bond with no recovery and p is the price of non-defaultable zero coupon bond.

Defaultable zero coupon bond with partial recovery at the default

The price of a defaultable zero coupon bond with partial recovery at the default is

$$
d_D(t,T) = d_0(t,T) + \mathrm{R_{EC}}Q(t,T), \tag{5.2.13}
$$

where

$$
Q(t,T) = \mathrm{E}_{\mathbb{Q}}\left(\mathrm{e}^{-\int_t^{\zeta} r(s)\mathrm{d}s} 1_{\{t<\zeta\le T\}} \mid \mathscr{F}_t \vee \mathscr{H}_t\right), \tag{5.2.14}
$$

which is the expected value of 1 paid at the time of the default at the time t. Now

$$
\begin{aligned}
Q(t,T) &= \frac{1_{\{\zeta>t\}}}{\mathbb{Q}(\zeta>t|\mathscr{F}_t)}\mathrm{E}_{\mathbb{Q}}\left(\mathrm{e}^{-\int_t^{\zeta} r(s)\mathrm{d}s} 1_{\{t<\zeta\le T\}} \mid \mathscr{F}_t\right) && (5.2.15)\\
&=1_{\{\zeta>t\}}\mathrm{e}^{\int_0^t \lambda(s)\mathrm{d}s}\mathrm{E}_{\mathbb{Q}}\left(\int_0^{\infty} 1_{\{t<\zeta\le T\}} D(t,s) 1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_t\right) && (5.2.16)\\
&=1_{\{\zeta>t\}}\mathrm{e}^{\int_0^t \lambda(s)\mathrm{d}s}\mathrm{E}_{\mathbb{Q}}\left(\int_t^{T} D(t,s) 1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_t\right). && (5.2.17)
\end{aligned}
$$

Now we can use Fubini's theorem to evaluate

$$\mathrm{E}_{\mathbb{Q}}\left(\int_t^T D(t,s)1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_t\right) \tag{5.2.18}$$

$$=\mathrm{E}_{\mathbb{Q}}\left(\mathrm{E}_{\mathbb{Q}}\left(\int_t^T D(t,s)1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_T\right) \mid \mathscr{F}_t\right) \tag{5.2.19}$$

$$=\mathrm{E}_{\mathbb{Q}}\left(\int_t^T D(t,s)\,\mathrm{E}_{\mathbb{Q}}\left(1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_T\right) \mid \mathscr{F}_t\right) \tag{5.2.20}$$

$$=\mathrm{E}_{\mathbb{Q}}\left(\int_t^T D(t,s)\mathbb{Q}\left(s\le\zeta<s+\mathrm{d}s \mid \mathscr{F}_T\right) \mid \mathscr{F}_t\right) \tag{5.2.21}$$

$$=\mathrm{E}_{\mathbb{Q}}\left(\int_t^T D(t,s)\lambda(s)\mathrm{e}^{-\int_0^s\lambda(u)\mathrm{d}u}\mathrm{d}s \mid \mathscr{F}_t\right) \tag{5.2.22}$$

$$=\mathrm{E}_{\mathbb{Q}}\left(\int_t^T \mathrm{e}^{-\int_t^s r(u)\mathrm{d}u}\lambda(s)\mathrm{e}^{-\int_0^s\lambda(u)\mathrm{d}u}\mathrm{d}s \mid \mathscr{F}_t\right). \tag{5.2.23}$$

Thus

$$Q(t,T) = 1_{\{\zeta>t\}}\mathrm{E}_{\mathbb{Q}}\left(\int_t^T \lambda(s)\mathrm{e}^{-\int_t^s(r(u)+\lambda(u))\mathrm{d}u}\mathrm{d}s \mid \mathscr{F}_t\right) \tag{5.2.24}$$

$$= 1_{\{\zeta>t\}}\int_t^T \mathrm{E}_{\mathbb{Q}}\left(\lambda(s)\mathrm{e}^{-\int_t^s(r(u)+\lambda(u))\mathrm{d}u} \mid \mathscr{F}_t\right)\mathrm{d}s \tag{5.2.25}$$

and

$$d_D(t,T) = d_0(t,T) + 1_{\{\zeta>t\}}\mathrm{R}_{\mathrm{EC}}\int_t^T \mathrm{E}_{\mathbb{Q}}\left(\lambda(s)\mathrm{e}^{-\int_t^s(r(u)+\lambda(u))\mathrm{d}u} \mid \mathscr{F}_t\right)\mathrm{d}s \tag{5.2.26}$$

5.2.1. The protection leg of a credit default swap

We now developed a price for the protection leg that pays L_{GD} at the default ζ, if $S<\zeta\le T$. The price of it at the time $0\le t<T$ is

$$\mathrm{CDS}_{\mathrm{pro}}(t) = 1_{\{\zeta>t\}}\mathrm{E}_{\mathbb{Q}}\left(1_{\{S<\zeta<T\}}D(t,\zeta)\mathrm{L}_{\mathrm{GD}} \mid \mathscr{G}_t\right) \tag{5.2.27}$$

$$= \frac{1_{\{\zeta>t\}}}{\mathbb{Q}(\zeta>t|\mathscr{F}_t)}\mathrm{E}_{\mathbb{Q}}\left(1_{\{S<\zeta<T\}}D(t,\zeta)\mathrm{L}_{\mathrm{GD}} \mid \mathscr{F}_t\right). \tag{5.2.28}$$

Now heuristically

$$
\begin{aligned}
&\mathrm{E}_{\mathbb{Q}}\left(1_{\{S<\zeta<T\}} D(t,\zeta) \mid \mathscr{F}_t\right) && (5.2.29)\\
=&\mathrm{E}_{\mathbb{Q}}\left(\int_0^\infty 1_{\{S<s<T\}} D(t,s) 1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_t\right) && (5.2.30)\\
=&\mathrm{E}_{\mathbb{Q}}\left(\int_S^T D(t,s) 1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_t\right) && (5.2.31)\\
=&\mathrm{E}_{\mathbb{Q}}\left(\mathrm{E}_{\mathbb{Q}}\left(\int_S^T D(t,s) 1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_T\right) \mid \mathscr{F}_t\right) && (5.2.32)\\
=&\mathrm{E}_{\mathbb{Q}}\left(\int_S^T D(t,s)\,\mathrm{E}_{\mathbb{Q}}\left(1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_T\right) \mid \mathscr{F}_t\right) && (5.2.33)\\
=&\mathrm{E}_{\mathbb{Q}}\left(\int_S^T D(t,s)\mathbb{Q}\left(s\le\zeta<s+\mathrm{d}s \mid \mathscr{F}_T\right) \mid \mathscr{F}_t\right) && (5.2.34)\\
=&\mathrm{E}_{\mathbb{Q}}\left(\int_S^T D(t,s)\lambda(u)\exp^{-\int_0^s \lambda(u)\mathrm{d}u}\mathrm{d}s \mid \mathscr{F}_t\right) && (5.2.35)\\
=&\mathrm{E}_{\mathbb{Q}}\left(\int_S^T \exp^{-\int_t^s r(u)\mathrm{d}u}\lambda(u)\exp^{-\int_0^s \lambda(u)\mathrm{d}u}\mathrm{d}s \mid \mathscr{F}_t\right) && (5.2.36)\\
=&\exp^{\int_0^t \lambda(u)\mathrm{d}u}\,\mathrm{E}_{\mathbb{Q}}\left(\int_S^T \lambda(s)\exp^{-\int_t^s (r(u)+\lambda(u))\mathrm{d}u}\mathrm{d}s \mid \mathscr{F}_t\right). && (5.2.37)
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathrm{CDS}_{\mathrm{pro}}(t) &= 1_{\{\zeta>t\}}\mathrm{L_{GD}}\,\mathrm{E}_{\mathbb{Q}}\left(\int_S^T \lambda(s)\exp^{-\int_t^s (r(u)+\lambda(u))\mathrm{d}u}\mathrm{d}s \mid \mathscr{F}_t\right) && (5.2.38)\\
&= 1_{\{\zeta>t\}}\mathrm{L_{GD}}\int_S^T \mathrm{E}_{\mathbb{Q}}\left(\lambda(u)\exp^{-\int_t^s (r(u)+\lambda(s))\mathrm{d}u} \mid \mathscr{F}_t\right)\mathrm{d}s, && (5.2.39)
\end{aligned}
$$

where we have assumed that the $\mathrm{L_{GD}}$ is a constant.

5.2.2. The premium leg of a credit default swap

The premium leg of a CDS with a coupon rate C has a value

$$
\begin{aligned}
\mathrm{CDS}_{\mathrm{pre}}(t,C) =&1_{\{\zeta>t\}}\,\mathrm{E}_{\mathbb{Q}}\left(D(t,\zeta)C^{h(\zeta)}1_{\{S<\zeta<T\}} \mid \mathscr{G}_t\right) && (5.2.40)\\
&+1_{\{\zeta>t\}}\sum_{i=1}^{n}\mathrm{E}_{\mathbb{Q}}\left(D(t,t_i)C_i 1_{\{\zeta>t_i\}} \mid \mathscr{G}_t\right), && (5.2.41)
\end{aligned}
$$

where $C_i = C\Delta(t_{i-1},t_i)$, $t_{h(\zeta)}$ is the last coupon date before the default (if it occurs) and $C^{h(\zeta)} = C\Delta(t_{h(\zeta)},\zeta) \approx C(\zeta - t_{h(\zeta)})$. We have also re-indexed the coupon dates so that $t_0 \le t \le t_1$.

Now

$$C_i(t,T) = \mathrm{E}_{\mathbb{Q}}\left(D(t,t_i)1_{\{\zeta>t_i\}} \mid \mathscr{G}_t\right) \tag{5.2.42}$$

$$= 1_{\{\zeta>t_i\}} \exp^{\int_0^t \lambda(s)\mathrm{d}s} \mathrm{E}_{\mathbb{Q}}\left(D(t,t_i)1_{\{\zeta>t_i\}} \mid \mathscr{F}_t\right) \tag{5.2.43}$$

$$= 1_{\{\zeta>t_i\}} \exp^{\int_0^t \lambda(s)\mathrm{d}s} \mathrm{E}_{\mathbb{Q}}\left(\exp^{\int_t^{t_i} r(s)\mathrm{d}s} \exp^{\int_0^{t_i} \lambda(s)\mathrm{d}s} \mid \mathscr{F}_t\right) \tag{5.2.44}$$

$$= 1_{\{\zeta>t_i\}} \mathrm{E}_{\mathbb{Q}}\left(\exp^{\int_t^{t_i} (r(s)-\lambda(s))\mathrm{d}s} \mid \mathscr{F}_t\right) \tag{5.2.45}$$

and, by recycling the earlier calculations, we get that

$$C^{h(\zeta)}(t,T) = \mathrm{E}_{\mathbb{Q}}\left((\zeta - t_{h(\zeta)})D(t,\zeta)1_{\{S<\zeta<T\}} \mid \mathscr{G}_t\right) \tag{5.2.46}$$

$$= 1_{\{\zeta>t\}} \exp^{\int_0^t \lambda(s)\mathrm{d}s} E \tag{5.2.47}$$

where the expectation

$$E = \mathrm{E}_{\mathbb{Q}}\left((\zeta - t_{h(\zeta)})D(t,\zeta)1_{\{S<\zeta<T\}} \mid \mathscr{F}_t\right) \tag{5.2.48}$$

$$= \mathrm{E}_{\mathbb{Q}}\left(\int_t^\infty (s - t_{h(s)})D(t,s)1_{\{S<s<T\}}1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_t\right) \tag{5.2.49}$$

$$= \mathrm{E}_{\mathbb{Q}}\left(\mathrm{E}_{\mathbb{Q}}\left(\int_S^T (s - t_{h(s)})D(t,\zeta)1_{\{s\le\zeta<s+\mathrm{d}s\}} \mid \mathscr{F}_T\right) \mid \mathscr{F}_t\right) \tag{5.2.50}$$

$$= \mathrm{E}_{\mathbb{Q}}\left(\int_S^T (s - t_{h(s)})D(t,s)\mathbb{Q}(s \le \zeta < s+\mathrm{d}s \mid \mathscr{F}_T) \mid \mathscr{F}_t\right) \tag{5.2.51}$$

$$= \mathrm{E}_{\mathbb{Q}}\left(\int_S^T (s - t_{h(s)})D(t,s)\mathrm{e}^{-\int_0^s \lambda(u)\mathrm{d}u}\lambda(s)\,\mathrm{d}s \mid \mathscr{F}_t\right) \tag{5.2.52}$$

$$= \int_S^T \mathrm{E}_{\mathbb{Q}}\left((s - t_{h(s)})D(t,s)\mathrm{e}^{-\int_0^s \lambda(u)\mathrm{d}u}\lambda(s) \mid \mathscr{F}_t\right)\mathrm{d}s. \tag{5.2.53}$$

Hence

$$A = CC^{h(\zeta)}(t,T) \tag{5.2.54}$$

$$= 1_{\{\zeta>t\}}C\int_S^T \mathrm{E}_{\mathbb{Q}}\left((s - t_{h(s)})D(t,s)\mathrm{e}^{-\int_t^s \lambda(u)\mathrm{d}u}\lambda(s) \mid \mathscr{F}_t\right)\mathrm{d}s. \tag{5.2.55}$$

Thus

$$\mathrm{CDS}_{\mathrm{pre}}(t,C) = 1_{\{\zeta>t\}}C\left(C^{s(\zeta)}(t,T) + \sum_{i=1}^n \Delta(t_{i-1},t_i)C_i(t,T)\right) \tag{5.2.56}$$

$$= C\left(\sum_{i=1}^n \Delta(t_{i-1},t_i)d_0(t,t_i)\right) + A, \tag{5.2.57}$$

where d_0 is the price of zero coupon bond with no recovery and A is term representing the accrued coupon before the default.

As we saw here, the most complicated part in the pricing of the premium leg of a CDS is the accrued coupon payment before the default. If we wish to simplify the model, then the accrual payment could be dropped or we could assume that premium is paid continuously. Both will result in biased priced, but this might be acceptable. If the accrued coupon payment is dropped, then the premium leg is just a portfolio of defaultable zero-coupon bonds with no recovery.

If the accrued coupon payment term has to simplified, it could be assumed that default happens in the middle of the coupon period or that the coupon is paid continuously during the accrual period.

Premium leg of a CDS with continuous premium

We may also suppose that premium leg pays a continuous premium c. If $\mathrm{d}t > 0$ is small, then premium leg pays from t to $\mathrm{d}t$ the amount of $c\mathrm{d}t$ assuming that the credit event does not occur. By using the old tricks, we may value this default intensity as

$$
\begin{aligned}
&\mathrm{E}_{\mathbb{Q}}\left(c\mathrm{e}^{-\int_t^{t+\mathrm{d}t} r(s)\mathrm{d}s}\mathrm{d}t 1_{\{\zeta>t+\mathrm{d}t\}} \mid \mathscr{G}_t\right) && (5.2.58)\\
=&1_{\{\zeta>t\}}\mathrm{E}_{\mathbb{Q}}\left(c\mathrm{e}^{-\int_t^{t+\mathrm{d}t}(r(s)+\lambda(s))\mathrm{d}s}\mathrm{d}t \mid \mathscr{F}_t\right). && (5.2.59)
\end{aligned}
$$

By taking the limit of this process we have that

$$
\begin{aligned}
\mathrm{CDS}_{\mathrm{pre}}(t,c) &= c\int_t^T 1_{\{\zeta>u\}}\mathrm{E}_{\mathbb{Q}}\left(\mathrm{e}^{-\int_t^u(r(s)+\lambda(s))\mathrm{d}s} \mid \mathscr{F}_t\right)\mathrm{d}u && (5.2.60)\\
&= c\int_t^T d_0(t,s)\mathrm{d}s, && (5.2.61)
\end{aligned}
$$

where $d_0(t,u)$ is the price of a defaultable u-bond with no recovery at the time t.

5.3. The assumption that the default is independent from interest rates

All the pricing formulas had the term

$$
\mathrm{E}_{\mathbb{Q}}\left(\exp^{-\int_t^T(r(s)+\lambda(s))\mathrm{d}s} \mid \mathscr{F}_t\right). \tag{5.3.1}
$$

If the default is independent of the interest rates under the risk neutral measure, then we may write

$$
\mathrm{E}_{\mathbb{Q}}\left(\exp^{-\int_t^T(r(s)+\lambda(s))\mathrm{d}s} \mid \mathscr{F}_t\right) = p(t,T)\Gamma(t,T), \tag{5.3.2}
$$

where

$$\Gamma(t,T) = \mathrm{E}_{\mathbb{Q}}\left(\exp^{-\int_t^T \lambda(s)\mathrm{d}s} \mid \mathscr{F}_t\right) \tag{5.3.3}$$

Hence

$$d_0(t,T) = 1_{\{\zeta>t\}} p(t,T)\Gamma(t,T) \tag{5.3.4}$$

$$d_M(t,T) = 1_{\{\zeta>t\}} p(t,T)\left(\Gamma(t,T)\mathrm{L}_{\mathrm{GD}} + \mathrm{R}_{\mathrm{EC}}\right) \tag{5.3.5}$$

for bonds with zero recovery or partial recovery at the maturity. As now

$$\begin{aligned} Q(t,T) &= 1_{\{\zeta>t\}} \int_t^T \mathrm{E}_{\mathbb{Q}}\left(\lambda(s)\mathrm{e}^{-\int_t^s (r(u)+\lambda(u))\mathrm{d}u} \mid \mathscr{F}_t\right)\mathrm{d}s & (5.3.6)\\ &= 1_{\{\zeta>t\}} \int_t^T p(t,s)\,\mathrm{E}_{\mathbb{Q}}\left(\lambda(s)\mathrm{e}^{-\int_t^s \lambda(u)\mathrm{d}u} \mid \mathscr{F}_t\right)\mathrm{d}s & (5.3.7)\\ &= 1_{\{\zeta>t\}} \int_t^T p(t,s)\,\mathrm{E}_{\mathbb{Q}}\left(-\frac{\partial}{\partial s}\mathrm{e}^{-\int_t^s \lambda(u)\mathrm{d}u} \mid \mathscr{F}_t\right)\mathrm{d}s, & (5.3.8) \end{aligned}$$

we have that

$$d_D(t,T) = d_0(t,T) + 1_{\{\zeta>t\}}\mathrm{R}_{\mathrm{EC}} \int_t^T p(t,s)\,\mathrm{E}_{\mathbb{Q}}\left(-\frac{\partial}{\partial s}\mathrm{e}^{-\int_t^s \lambda(u)\mathrm{d}u} \mid \mathscr{F}_t\right)\mathrm{d}s \tag{5.3.9}$$

5.4. $A(M,N)$ model for credit risk

We assume that there are N state-variables driving short-rate and intensity processes under the risk-neutral measures. Of these, M follow square-root process and $N-M$ are gaussian. We follow the presentation in Nawalka et al. (2007, pp. 457–476).

More precisely, the correlated gaussian process are

$$\mathrm{d}Y_i(t) = -k_i Y_i(t)\mathrm{d}t + \nu_i \mathrm{d}W_i(t), \tag{5.4.1}$$

where W_i is a Wiener process and

$$\mathrm{d}W_i(t)\mathrm{d}W_j(t) = \rho_{ij}\mathrm{d}t \tag{5.4.2}$$

for all $i,j = 1,2,\dots,N-M$. Here $-1 < \rho_{ij} = \rho_{ji} < 1$ and $\rho_{ii} = 1$. The M square-root processes are

$$\mathrm{d}X_m(t) = \alpha_m(\theta_m - X_m(t))\mathrm{d}t + \sigma_m\sqrt{X_m(t)}\mathrm{d}Z_m(t) \tag{5.4.3}$$

where Z_m are independent Wiener process and

$$\mathrm{d}W_i(t)\mathrm{d}Z_m(t) = 0 \tag{5.4.4}$$

for all $i = 1, 2, \ldots, N-M$ and $m = 1, 2, \ldots, M$. The short-rate is defined by

$$r(t) = \delta_r + \sum_{m=1}^{M} a_m X_m(t) + \sum_{i=1}^{N-M} c_i Y_i(t), \tag{5.4.5}$$

and the default intensity by

$$\lambda(t) = \delta_\lambda + \sum_{m=1}^{M} b_m X_m(t) + \sum_{i=1}^{N-M} d_i Y_i(t), \tag{5.4.6}$$

where δ_r, δ_λ are constants and a_m, c_m for $m = 1, 2, \ldots, M$ and c_i, d_i for $i = 1, 2, \ldots, N-M$ for non-negative constants.

As shown earlier in Equation 5.2.12, the price of a defaultable T-bond with partial recovery of a face value at the maturity is given by

$$d(t,T) = d_0(t,T) + \mathrm{R_{EC}} \int_t^T \mathrm{E}_{\mathbb{Q}} \left(\lambda(u) \mathrm{e}^{-\int_t^u (r(s)+\lambda(s))\mathrm{d}s} \mid \mathscr{F}_t \right) \mathrm{d}u, \tag{5.4.7}$$

where

$$d_0(t,T) = \mathrm{E}_{\mathbb{Q}} \left(\mathrm{e}^{-\int_t^T (r(s)+\lambda(s))\mathrm{d}s} \mid \mathscr{F}_t \right) \tag{5.4.8}$$

is the price of a defaultable bond with no recovery.

We denote

$$G(t,T) = \mathrm{E}_{\mathbb{Q}} \left(\lambda(T) \mathrm{e}^{-\int_t^T (r(s)+\lambda(s))\mathrm{d}s} \mid \mathscr{F}_t \right) \tag{5.4.9}$$

$$= \frac{\partial}{\partial \phi} \left(\eta(t,T,\phi) \right)_{\phi=0}, \tag{5.4.10}$$

where

$$\eta(t,T,\phi) = \mathrm{E}_{\mathbb{Q}} \left(\mathrm{e}^{-\int_t^T (r(s)+\lambda(s))\mathrm{d}s} \mathrm{e}^{\phi\lambda(T)} \mid \mathscr{F}_t \right) \tag{5.4.11}$$

The solution to this expectation is given by (under certain assumptions)

$$\eta(t,T,\phi) = \mathrm{e}^{A^\dagger(\tau) - \sum_{m=1}^{M} (a_m+b_m) B_m^\dagger(\tau) X_m(t) - \sum_{i=1}^{N-M} (c_i+d_i) C_i^\dagger(\tau) Y_i(t) - H^\dagger(t,T)}, \tag{5.4.12}$$

where $\tau = T - t$ and

$$H^{\dagger}(t,T) = \int_t^T (\delta_r + \delta_\lambda)\mathrm{d}x = (\delta_r + \delta_\lambda)\tau, \tag{5.4.13}$$

$$\beta_{1m} = \sqrt{\alpha_m^2 + 2(a_m + b_m)\sigma_m^2}, \tag{5.4.14}$$

$$\beta_{2m} = \frac{-\alpha_m + \beta_{1m}}{2}, \tag{5.4.15}$$

$$\beta_{3m} = \frac{-\alpha_m - \beta_{1m}}{2}, \tag{5.4.16}$$

$$\beta_{4m} = \frac{-\alpha_m - \beta_{1m} + \phi b_m \sigma_m^2}{-\alpha_m + \beta_{1m} + \phi b_m \sigma_m^2}, \tag{5.4.17}$$

$$B_m^{\dagger}(\tau) = \frac{2}{(a_m + b_m)\sigma_m^2}\left(\frac{\beta_{2m}\beta_{4m}\mathrm{e}^{\beta_{1m}\tau} - \beta_{3m}}{\beta_{4m}\mathrm{e}^{\beta_{1m}\tau} - 1}\right), \tag{5.4.18}$$

$$A_X^{\dagger}(\tau) = \sum_{m=1}^{M} \frac{\alpha_m \theta_m}{\sigma_m^2}\left(\beta_{3m}\tau + \log\left(\frac{1 - \beta_{4m}\mathrm{e}^{\beta_{1m}\tau}}{1 - \beta_{4m}}\right)\right) \tag{5.4.19}$$

$$q_i = 1 + \phi k_i \frac{d_i}{c_i + d_i} \tag{5.4.20}$$

$$C_i^{\dagger}(\tau) = \frac{1 - q_i \mathrm{e}^{-k_i \tau}}{k_i}, \tag{5.4.21}$$

$$D^{\dagger}(\tau) = \tau - q_i C_i^{\dagger}(\tau) - q_j C_j^{\dagger}(\tau) + q_i q_j \frac{1 - \mathrm{e}^{-(k_i + k_j)\tau}}{k_i + k_j} \tag{5.4.22}$$

$$A_Y^{\dagger}(\tau) = \sum_{i=1}^{N-M}\sum_{j=1}^{N-M} \frac{(c_i + d_i)(c_j + d_j)\nu_i \nu_j \rho_{ij}}{k_i k_j} D^{\dagger}(\tau) \tag{5.4.23}$$

$$A^{\dagger}(\tau) = \phi\delta_\lambda + \frac{1}{2}A_Y^{\dagger}(\tau) - 2A_X^{\dagger}(\tau) \tag{5.4.24}$$

for all $i = 1, 2, \ldots, N - M$ and $m = 1, 2, \ldots, M$. Now $G(t,T)$ can be approximated as the numerical derivative of $\eta(t,T,\phi)$ at 0.

Under this model

$$d_0(t,T) = \mathrm{e}^{A^{\dagger}(\tau) - \sum_{m=1}^{M}(a_m + b_m)B_m^{\dagger}(\tau)X_m(t) - \sum_{i=1}^{N-M}(c_i + d_i)C_i^{\dagger}(\tau)Y_i(t) - H^{\dagger}(t,T)} \tag{5.4.25}$$

, where $\tau = T - t$ and

$$H(t,T) = \int_t^T (\delta_r + \delta_\lambda)\,\mathrm{d}x = (\delta_r + \delta_\lambda)\,\tau, \tag{5.4.26}$$

$$\beta_m = \sqrt{\alpha_m^2 + 2(a_c + b_m)\sigma_m^2}, \tag{5.4.27}$$

$$B_m(\tau) = \frac{2(\mathrm{e}^{\beta_m \tau} - 1)}{\beta_m + (\mathrm{e}^{\beta_m \tau} - 1) + 2\beta_m}, \tag{5.4.28}$$

$$A_X(\tau) = \sum_{m=1}^{M} \frac{\alpha_m \theta_m}{\sigma_m^2} \log \frac{2\beta_m \mathrm{e}^{\frac{(\beta_m + \alpha_m)\tau}{2}}}{(\beta_m + \alpha_m) + (\mathrm{e}^{\beta_m \tau} - 1) + 2\beta_m}, \tag{5.4.29}$$

$$C_i(\tau) = \frac{1 - \mathrm{e}^{-k_i \tau}}{k_i}, \tag{5.4.30}$$

$$D(\tau) = \tau - C_i(\tau) - C_j(\tau) + \frac{1 - \mathrm{e}^{-(k_i + k_j)\tau}}{k_i + k_j}, \tag{5.4.31}$$

$$A_Y(\tau) = \sum_{i=1}^{N_M} \sum_{j=1}^{N_M} \frac{(c_i + d_i)(c_j + d_j)\nu_i \nu_j \rho_{ij}}{k_i k_j} D(\tau), \tag{5.4.32}$$

$$A(\tau) = 2A_X(\tau) + \frac{1}{2} A_Y(\tau). \tag{5.4.33}$$

The risk-free bond $p(t,T)$ may be priced using the equation above, but with $b_m = 0$ and $d_i = 0$ for all $m = 1, 2, \ldots M$ and $i = 1, 2, \ldots, N - M$. By differentiating of $Y_i^*(t) = c_i Y_i(t)$, we get that

$$\begin{aligned}
\mathrm{d}Y_i^*(t) &= c_i \mathrm{d}Y_i(t) && (5.4.34)\\
&= -k_i c_i Y_i(t) + c_i \nu_i \mathrm{d}W_i(t) && (5.4.35)\\
&= -k_i Y_i^*(t) + (c_i \nu_i)\,\mathrm{d}W_i(t). && (5.4.36)
\end{aligned}$$

This implies that we may use the formulas in Subsection 4.3.1 by replacing ν_i with $c_i \nu_i$ and $Y_i(t)$ with $c_i Y_i(t)$. Similarly differentiating $X_m^* = a_m X_m(t)$ yields

$$\begin{aligned}
\mathrm{d}X_m^*(t) &= a_m \mathrm{d}X_m(t) && (5.4.37)\\
&= \alpha_m (a_m \theta_m - a_m X_m(t))\mathrm{d}t + \sigma_m \sqrt{a_m} \sqrt{a_m X_m(t)} \mathrm{d}Z_m(t) && (5.4.38)\\
&= \alpha_m ((a_m \theta_m) - X_m^*(t))\mathrm{d}t + (\sigma_m \sqrt{a_m}) \sqrt{X_m^*(t)} \mathrm{d}Z_m(t) && (5.4.39)
\end{aligned}$$

end we see that θ_m needs to be replaced to $a_m \theta_m$, σ_m to $\sigma_m \sqrt{a_m}$ and $X_m(t)$ to $a_m X_m(t)$. Thus we may also use the machinery of Chapter 4.5 to price derivatives that only depends on the risk-free rate with similar changes.

Since common state-variables may drive both risk-free rate and the default intensity, they may be correlated under the models of this family. As some of the state-variables may not be shared, these models have potentially a very rich structure.

We adopt the following notation. The model $D((a_X, b_X, c_X), (a_Y, b_Y, c_Y))$ is the model defined in the Equations 5.4.5 and 5.4.6 with the following properties

- a_X is the number of square-root processes that are present in both Equations 5.4.5 and 5.4.6,
- b_X is the number of square-root processes that are unique to the risk-free rate process,
- c_X is the number of square-root processes that are unique to the spread process,
- a_Y is the number of gaussian processes that are present in both Equations 5.4.5 and 5.4.6,
- b_Y is the number of gaussian processes that are unique to the risk-free rate process and
- c_Y is the number of gaussian processes that are unique to the spread process.

Thus it is $A(M,N)$ model with

$$M = a_X + b_X + c_X \tag{5.4.40}$$

$$N - M = a_Y + b_Y + c_Y. \tag{5.4.41}$$

We attempt to calibrate these models to market data in empirical work in section 6.2.2.

6. EMPIRICAL WORK

6.1. Research environment

The computing was done on home PC with i5 3.2 GHz CPU. The code was written in Python 3.6. The main workhorse was Scipy, which is a "Python-based ecosystem of open-source software for mathematics, science, and engineering (Jones, Oliphant, Peterson, et al. 2001–)". No serious effort was done to write computationally efficient code. The resulting algorithms are very sloppy, and due to hurry, are in dire need of some refactoring. The code did not try to utilize multi-cores, so most of the available computing power was not utilized.

By using the available data, the prices of several theoretical zero coupon bonds were calculated. The maturities of these instruments were taken directly from the retrieved data, so no bootstrapping or interpolation was done by the author.

The model parameters were chosen so that the sum of squared relative pricing errors were to be minimized. This minimization was done in two stages. First, an initial guess for solution was searched by using a home-made variant of differential evolution. This value was given as the initial value to the L-BFGS-B algorithm (see Byrd, Lu, Nocedal, and Zhu (1995) and Zhu, Byrd, Lu, and Nocedal (1997)) which has been implemented in SciPy. L-BFGS-B algorithm is a quasi-newtonian method so it uses an approximation of Jacobian matrix to guide iteration toward a local minimum point. It is well suited for optimization problems with large number of parameters but the performance depends very much on the quality of the initial guess as it will not do up hill climbing. However, since the curse of the dimensionality and the rather small population sizes (512 or 1024 for the initial populations), there is no guarantee that initial guesses were close to global minima.

All algorithms either uses authors own code or Scipy's standard libraries with one exception. The symmetry of the parameterized correlation matrix can be guaranteed trivially. But it has also be also positively semi-definite (see, for example, Higham (2002)). In order to to guarantee that the estimated matrix will actually be a valid correlation matrix, the code utilizes Python code by Croucher (2014–) which is an implementation of MATLAB code by Higham (2013).

The code, along with Jupyter notebooks used in data analysis, can by found at `https://github.com/mrytty/gradu-public` (Rytty 2019).

6.2. Stationary calibration

Data

The data that was used in stationary calibration part was gathered mainly from Eikon Datastream. It consists of 5 data sets:

- Interest rates derived from overnight indexed swap curve (OIS).
- Interest rates derived from swap curve.
- Yields derived from prices of Germany Government bonds.
- Yields derived from prices of France Government bonds.
- Yields derived from prices of Italy Government bonds.

The date for all these data sets is $7/26/2018$. Maturities for non-government rates ranged from overnight rate to 30-year rate and maturities for government rates ranged from 6-month rate to 30-year rate. Graphical presentation of implied interest rates, forward rates and zero-coupon bond prices can be seen in Figures B.3 and B.1. Interpolation of the missing rates for graphical purposes is done using either by linear interpolation, quadratic interpolation or cubic spline interpolation.

As we can see in Figure B.2, Italy is clear outlier. It has all positive rates, it's rates are much higher and the overall curve has a traditional" shape. Germany and France has pretty similar curves, but the spread between widens over time. Swap curve has an anomaly, for maturities less than one month, it has an odd hook which causes it be smaller than OIS curve. This should happen in the current paradigm, especially for the shortest maturities. There could have been an market anomaly or different curves might have distinct bootstrapping algorithms.

It should be noted that for IOS and swap curves, almost half of the data points are with a maturity less than a year. Therefore the calibration tends to weights the fitting in this section heavily.

6.2.1. Models without credit risk

Calibration was done for 12 models:

- 1-factor: $A(0,1)+$, $A(1,1)+$

- 2-factor: $A(0,2)+$, $A(1,2)+$, $A(2,2)+$
- 3-factor: $A(0,3)+$, $A(1,3)+$, $A(2,3)+$, $A(3,3)+$
- 4-factor: $A(2,4)+$, $A(3,4)+$
- 5-factor: $A(4,5)+$

These are all of the type $A(M,N)+$. A meaningful calibration of free parameters in $A(M,N)++$ -models requires calibration to cap or swaption prices. Hence they need repeated calculations of bond options prices. For majority of multi-factor models we have no explicit analytical pricing formula[1]. Since the Fourier transformation method requires the numerical integration, which is computationally costly without FFT, calibrations of $A(M,N)++$ -models was not attempted. For this and due to lack of data, no caps or swaptions data were used in model calibration for simple $+$-models. As the analytical option prices for zero coupon bonds exists for $A(0,1)$, $A(1,1)$ and $A(0,2)$ models, calibration for respective $++$-models could have been done using these formulas.

The curse of dimensionality refers to the mathematical fact that the sparseness of the optimization space grows exponentially when the dimension of the problem increases. For example, n-dimensional hypercube has 2^n vertexes. But even if we double the number of sample points with each dimension, the average Euclidean distance between the points keeps growing as the longest diagonal of a n-dimensional hypercube is $\sqrt{n}$.

These affine models have a large number of free parameters. A square root process has 4 parameters, a Gaussian process has 3 free parameters and n Gaussian processes need $\frac{n(n-1)}{2}$ more parameters for correlations. Also the combined model need one additional parameter for the shift. As we can see from the Table 6.1, the parameter spaces for the multi-factor models are rather large. As $1024 = 2^{10}$, we see that the initial population used in the differential evolution is much smaller than the number of vertices in corresponding hypercube for 3-factor models.

[1] Single factor models and $G2++$ are exceptions as we have seen.

Model	Parameters
$A(0,1)+$	4
$A(1,1)+$	5
$A(0,2)+$	8
$A(1,2)+$	8
$A(2,2)+$	9
$A(0,3)+$	13
$A(1,3)+$	12
$A(2,3)+$	12
$A(3,3)+$	13
$A(3,4)+$	16
$A(4,5)+$	20

Table 6.1: Number of parameters per model

We note that IOS and swap data set has 13 points, France has 8 and Germany and Italy has only 7 data points. For many of these calibrations this means that there are fewer data points than parameters. However, this is not as serious problem as under-determined fitting in linear regressions. The reason is that not every zero curve can be replicated in $A(M,N)+$-models. If the model can't fit a curve with certain number of points, then adding an additional point will not make the fit any better. On the contrary, adding an additional interpolated point will probably make the fit of original points worse[2] Theoretically the ill-poised calibration might be a problem but by the results we get, it does not seem to be so.

A differential evolution (DE) algorithm with starting population of 1024 and minimum population of 32 runnning for 1000 generations is guaranteed to evaluate only about 33000 samples. In practice there seemed to be at least 40000 pricing function calls which is still insignificant sample from the optimization space for majority of these models. As this kind of optimization landscape probably has lots of local minimums, therefore it is quite likely that the DE does not explore optimization space enough with these computational resources. This could be solved by using larger populations or by using algorithm that tend to explore the space more efficiently. A potential candidate could have been a variant of particle swarm optimization algorithms that were introduced by Eberhart and Kennedy (1995). In practice, these kind of problems are often solved with another stochastic optimization method called simulated annealing algorithms (SA). One major difference in SA and DE algorithms is that in SA maintains only one candidate solution which is varied. SA is thus better in problems, where the initial guess will be made reasonable well.

There is also the problem of symmetries for certain models. For example, the mod-

[2]Since we are dealing with stochastic algorithm, it could also make the fit better.

els $A(0,2)+$ and $G2++$ are symmetric regarding the Gaussian factors. So if the optimization function has at least one global minima, then it has also another global minima that is achieved by flipping the Gaussian factors. In theory, this could conversely affect convergence as DE algorithm will have sub-populations converging to different optimum values.

Results

All following errors are expressed as relative pricing differences in percentage points.

The best parameter sets per model and curve are presented in Tables B.1, B.2, B.3, B.4 and B.5. There are lots of cases that the parameter is at the border of the optimization space. For example, correlation coefficients were contained to a line segment $[-0.99.0.99]$. Several models with correlations show ρ close to -1 or 1. This implies that the extra factor is not actually present but those highly correlated factors change in unison. Other example of border cases are diffusion parameters that had lower bounds of 0.001.

The resulting errors by maturity are shown in Figure 6.1. The charts on left show relative pricing errors and the charts on right show absolute value of relative error. We see that these models calibrations tend to give very similar results with some exceptions. Almost every model and every curve tend to underprice 10-year maturities. OIS and swap curves are overpriced at 20-year maturities underpriced at 30-year maturities. For German and French curve 20-year maturity is underpriced but for Italy it is overpriced. Overall the fitting quality is poor, because there are plenty of errors with magnitudes over half a percent. Italian curve seems to be easier curve to fit than the others and the model $A(3,4)+$ seems to produce significantly better fit than others for OIS and swap curves.

Figure 6.1: Calibration errors for models without credit risk

Figures B.5–B.16 show how the models struggle in rate perspective to describe the early maturities. However, as the small time-factor lessens errors in zero-coupon prices, these errors are not significant in price perspective.

Overall, the errors are highly correlated as seen in the Figure 6.2. Especially multi-factor models have high error correlations and tend to produce similar results.

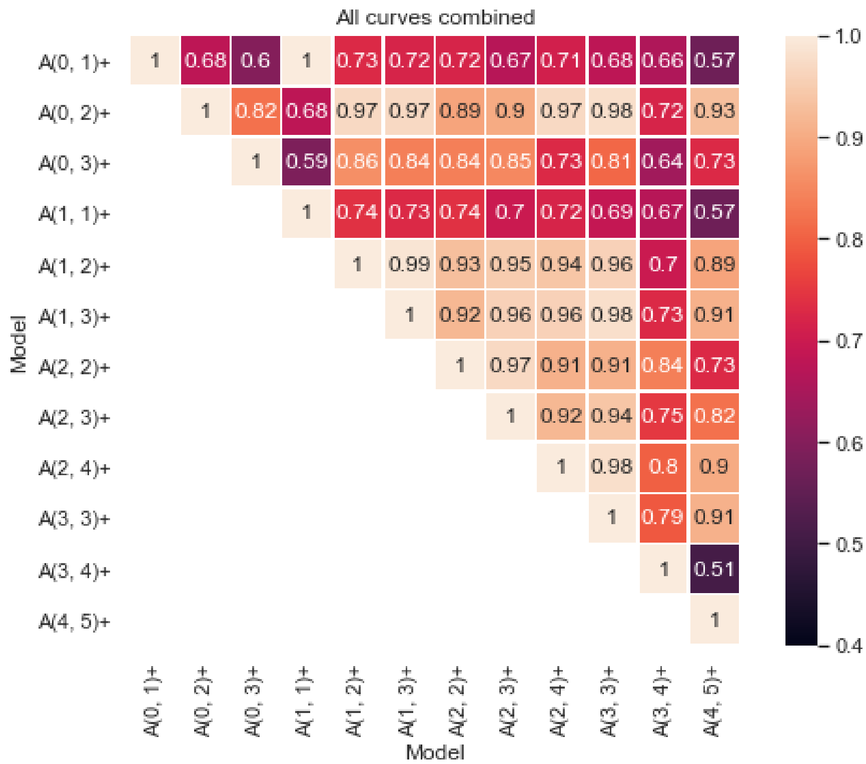


Figure 6.2: Error correlations for models without credit risk

In order to determine the significance of randomness in algorithm, a second calibration with different random number generator seeds was done to selected models. For single factor models, the calibration errors had very little variation. This is presented in Figure B.17. The figures B.18 and B.19 shows comparison for multi-factor models. For these models, there are significantly more variation. Significantly the alternative calibration of $A(3,3)+$ fits swap rates well but does poorly on OIS rates. The comparison tables for different calibrations are shown in Appendix B.5. Single factor model $A(0,1)+$ shows almost identical parameters between curves but the model $A(1,1)+$ has some variation. For multi-factor models the resulting parameters have significant differences even when accounting for symmetries.

The resulting mean absolute errors are shown in Figure 6.3. The term-structure of Italy looks to be straight-forward as every calibration tends to be fit it quite well. One reason for the good fit of Italy might be the fact that is the only curve that resembles the pre-crisis curves. Some models seem to have a excellent performance with OIS and Swap structures while all the models seem to struggle with France.

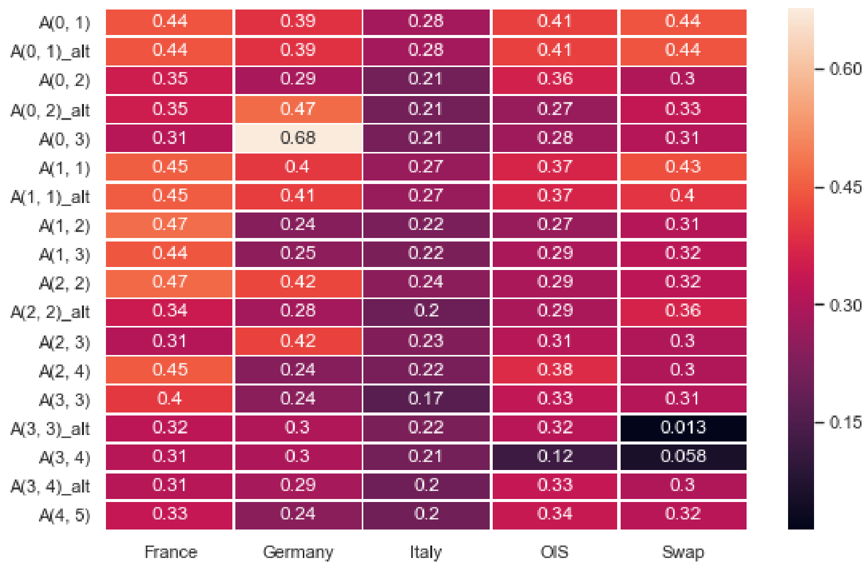


Figure 6.3: Mean absolute calibration error for models without credit risk

Again we see that there are no significant quality differences between single factor models. We also see that fitting quality seems to increase with the number of factors in the model, as is expected. The models with correlating Gaussian factors ($A(0,2)+$,$A(0,3)+$,$A(1,3)+$,$A(2,4)+$) do not seem perform well. In this sample, the models $A(3,3)+$ and $A(3,4)+$ tend to have the best performance, especially for OIS and Swap curves.

Due to several sources of uncertainty, it is hard to draw any definitive conclusions from these results. First, no instruments whose prices depend heavily on rate volatility were used. Secondly, we have demonstrated that the given computational resources, the used calibration algorithm tends not perform consistently for multi-factor models. We note that Italy seem to be an easier curve to fit. One reason for this might be the fact that is the only curve that resembles the curves of the pre-crisis curves with all positive rates.

For single-factor models, the fitting quality is not good with the exception of Italy. For multi-factor models, the used calibration methods have been flawed to be used consistently. We saw some acceptable calibrations, which then could not be reproduced with different random number generate set ups. Therefore some of these multi-factor models could be used as references in model risk management practices. But this would warrant more serious effort to set up a proper optimization methodology.

Cap pricing comparison

In order to gauge the differences in implied volatility structure, we price three different caps using the calibrated OIS and swap models. For OIS models we used the $A(3,4)+$ as the reference model and for swap models it was $A(3,3)+$. The relative pricing errors are in Figures 6.4 and 6.4.

As we can see, the prices are widely different. In order to price caps, calibration needs to include instruments whose prices depend on the volatilities or the results will be garbage.

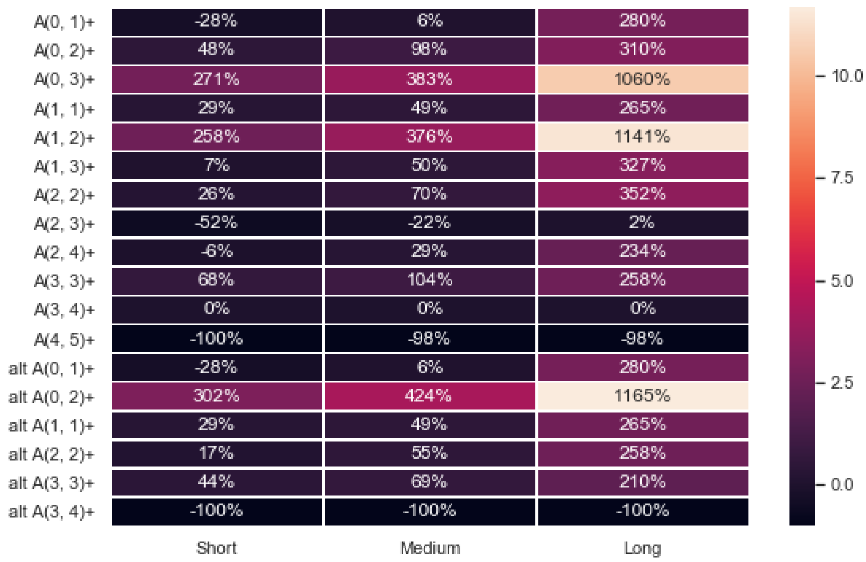


Figure 6.4: Relative pricing errors of sample caps compared to prices given by OIS $A(3,4)+$-model

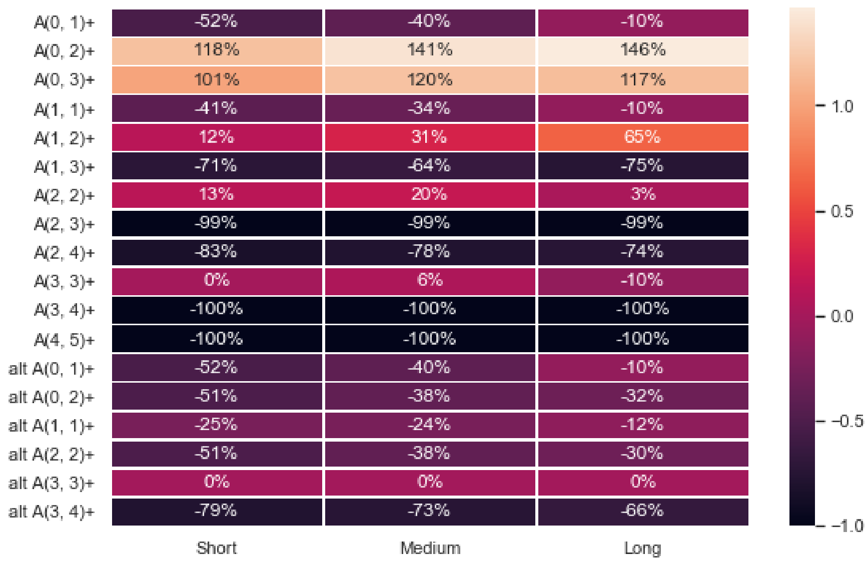


Figure 6.5: Relative pricing errors of sample caps compared to prices given by swap $A(3,3)+$-model

6.2.2. Models with credit risk

For default risk model, we considered four set-ups. They were

- OIS as a risk-free rate and swap rate as a risky rate[3],
- OIS as a risk-free rate and the rate for Italy as a risky rate,
- German rate as a risk-free rate and the rate for France as a risky rate and
- German rate as a risk-free rate and the rate for Italy as a risky rate,

Here we do not consider the yield for German Republic is a literal risk-free rate but rather we want to see how we can model the spread between it and the riskier rate. This will lead to bias but the effect is probably negligible.

Six affine models were considered and only 2-factor models for calibrated because of the performance considerations. The models were

- $D((0,0,0),(0,1,1))$,
- $D((0,0,0),(1,1,0))$,
- $D((0,0,1),(0,1,0))$,
- $D((0,0,1),(1,0,0))$,

[3]This is interesting case since the spread changes signs.

- $D((1,0,0),(0,1,0))$ and
- $D((1,0,0),(1,0,0))$.

Since both the risk-free rate and the risky were calibrated at the same time, the calibration function was heavier to calculate as it has twice as many instrument prices to. Thus we used a starting population of 512 and a mixumum of 500 generations.

In the calibration, LGD was not set, it was a free parameter to be optimized. Since the calculation for LGD residual demanded numerical integration which is resource heavy. In order to speed DE algorithm, LGD was set to 1 for DE search and it was not computed. Only the L-BFGS-B optimization had LGD as a free parameter.

The model $D((0,0,1),(0,1,0))$ is the only model where there are no interaction between the spread and interest rate process. It is also one of the best fitting models. However, due to deficiencies of the calibration algorithm, this does not necessary imply that spread and interest rate has very little interaction. It is very likely that the capturing both rate curve and differently shaped spread curve with only two-factors is impossible.

Results

The mean absolute errors are shown in Table 6.6. For unknown reasons, the optimizer got stuck when running $D((1,0,0),(0,1,0))$ model for the pair of Germany and France, so that model has only 3 calibrations. The calibrated parameters can be found in Appendix B.8.

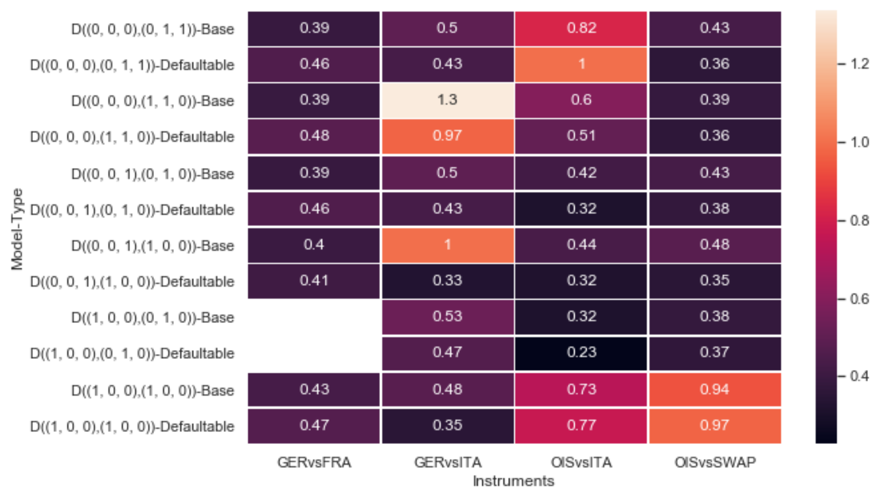


Figure 6.6: Mean absolute calibration error for models with default risk

Figures B.24 and B.25 show relative calibration errors. Since the spread between the German and French curves is minimal, it is not surprising that there are no significant differences between model performance in that case. Other cases show more variation between the models. Even for the best models, the worst relative pricing errors are close to one percent, which is unacceptable inaccuracy for practical purposes. It seems that two factors are not enough to capture both the interest rate and spread curves in this environment.

As we can see, the models do not perform well. The mean relative errors are around 0.5%. The models $D((0,0,1),(0,1,0))$ and $D(1,0,0),(0,1,0))$ have the best fits. However, these calibrations include very high LGD parameters and few corner values. For purely credit spread modeling, the model $D((0,0,1),(1,0,0))$ does fairly well.

However, due to deficiencies of the optimization methodology , these results should not be taken seriously.

6.3. Dynamic Euribor calibration without credit risk

Data

The data that was used in stationary calibration part was gathered from Eikon Datastream. It consisted of monthy Euribor rates from February 26, 2004 to January 26, 2019. The data had one week, two week, one month, three month, six month and one year rates. Thus it has $6*180 = 1080$ datapoints. The graphical presentation of the date can be found in Figures 6.7 and 6.7. The correlation of between different maturities is very high but changes have more variation as can be seen in Figure 6.9

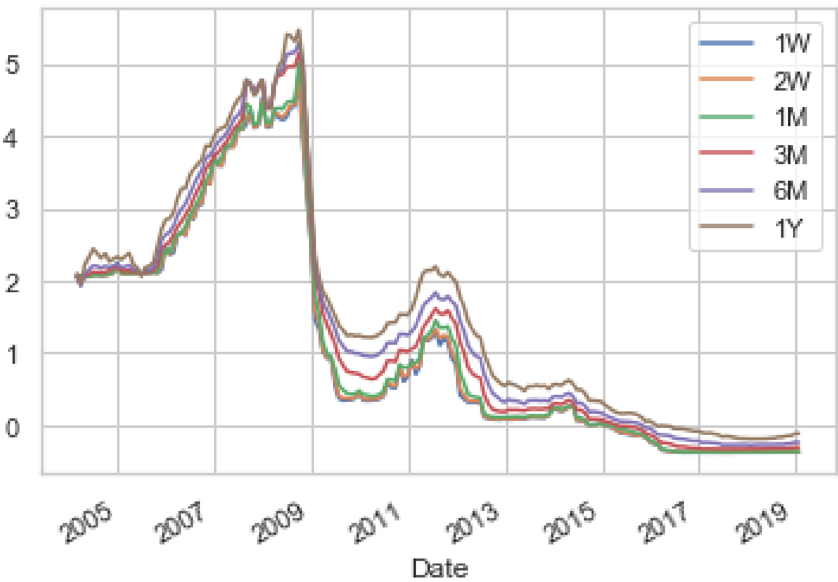


Figure 6.7: Euribor rates from February 26, 2004 to January 26, 2019

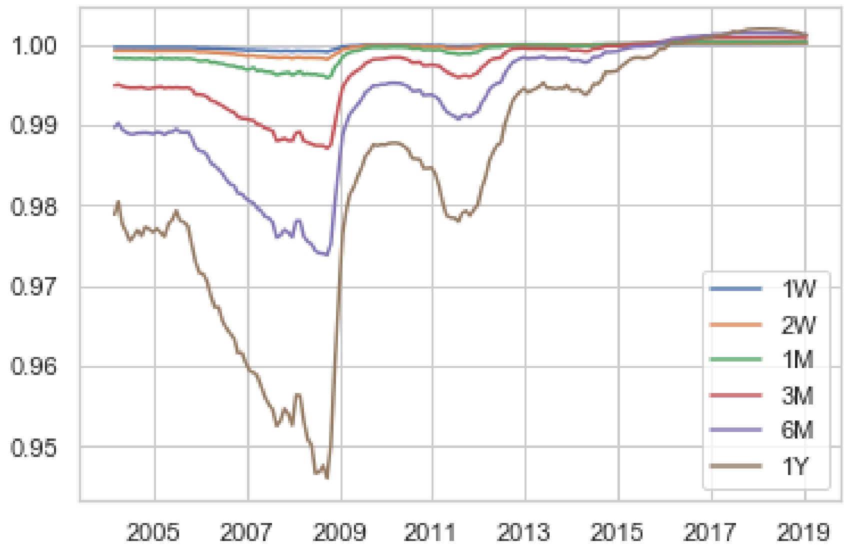


Figure 6.8: Implied Euribor discount factors from February 26, 2004 to January 26, 2019

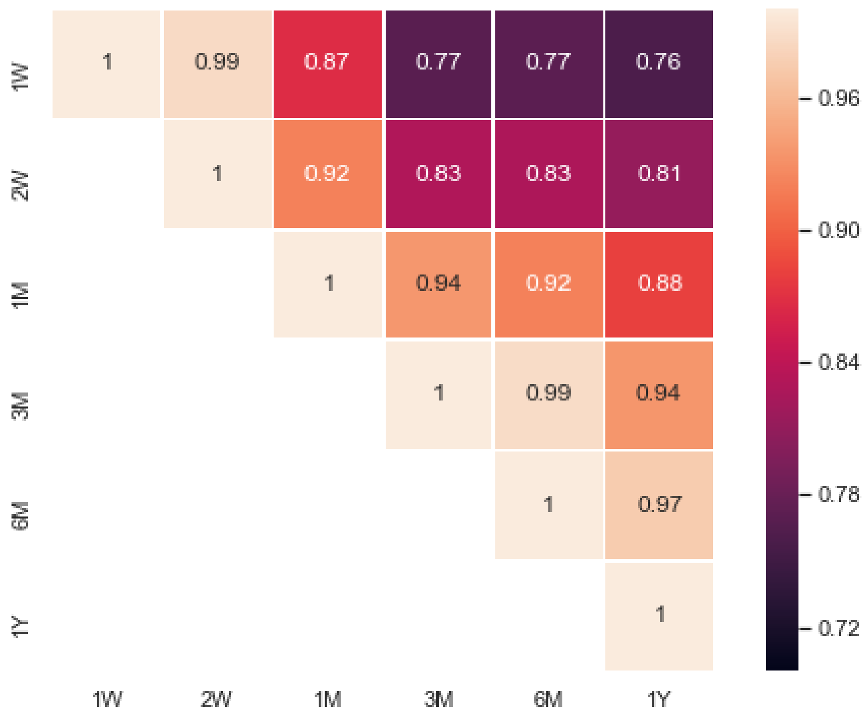


Figure 6.9: Correlation of Euribor rate changes from February 26, 2004 to January 26, 2019

The data is very varied as it has the rising rates of pre-crisis of 2007-08, the crisis period with sharply falling rates and the following tapering toward negative rates during the quantitative easing. The different shapes of the curves can be observed in Figure 6.10.

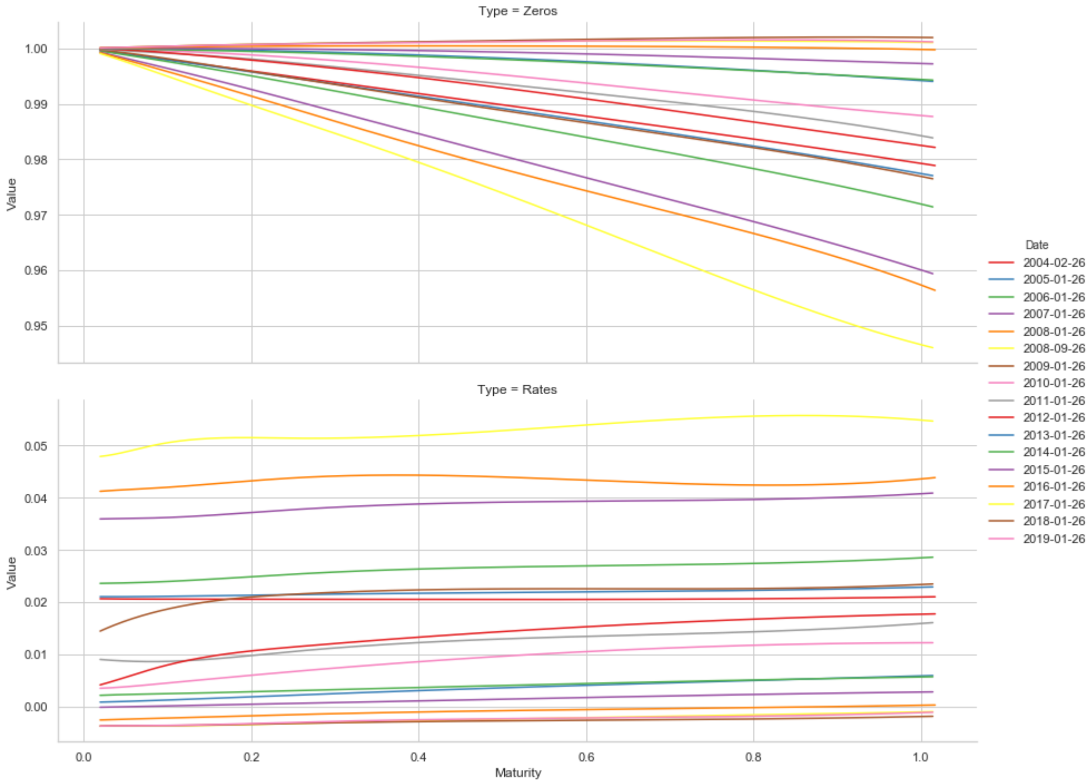


Figure 6.10: Euribor rate and discount curves

Calibration

The calibration was done for 5 different models: $A(0,1)+$, $A(1,1)+$, $A(0,2)+$, $A(1,2)+$ and $A(2,2)+$. The model was first calibrated for the very first date using differential evolution but for the next date calibration was done just by used L-BFGS-B algorithm with the initial parameters value. The assumption behind the choice was that the subsequent curves should be similar and therefore the parameters should be locally stable. But since the L-BFGS-B algorithm may not escape local minimums could also cause unnecessary stability in parameters. Parameter time series is presented in Figure 6.11.

Model = A(0, 1)+
Model = A(1, 1)+
Model = A(0, 2)+
Model = A(1, 2)+
Model = A(2, 2)+
Calibration
Date
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
2005
2007
2009
2011
2013
2015
2017
2019
Variable
Y_0
k
nu
shift
X_0
alpha
sigma
theta
Y_01
Y_02
k1
k2
nu1
nu2
rho
X_01
X_02
alpha1
alpha2
sigma1
sigma2
theta1
theta2

There is a curious observation conserning model$A(2,2)+$. The components of mean reversion speed α have high positive correlation but after the crisis the correlation is highly negative until after the 2011 the value stabilizes. This is also the the time when rates are getting very close to the zero so the stabilization is not surprising. This stabilization also can be seen in other models too, most notable in $A(0,2)+$ model.

The quality of the fitting was not very good. As we can see in Figure 6.12, every tested model had did not perform well during the crisis. Although our calibration algorithm could theoretically prohibit changes in parameter movements, this seems not be the case. If we compare Figures 6.11 and 6.12, we see that largest pricing errors occur during the times of volatility in parameter values. After 2013, all the models show very stable pricing errors and parameters. This is highly logical as then the rates are negative and discounting curve is very flat (for the maturities under a year). Errors by maturities are presented in Figures B.26. The biggest errors occurs in 1, 3 and 6 month maturities.

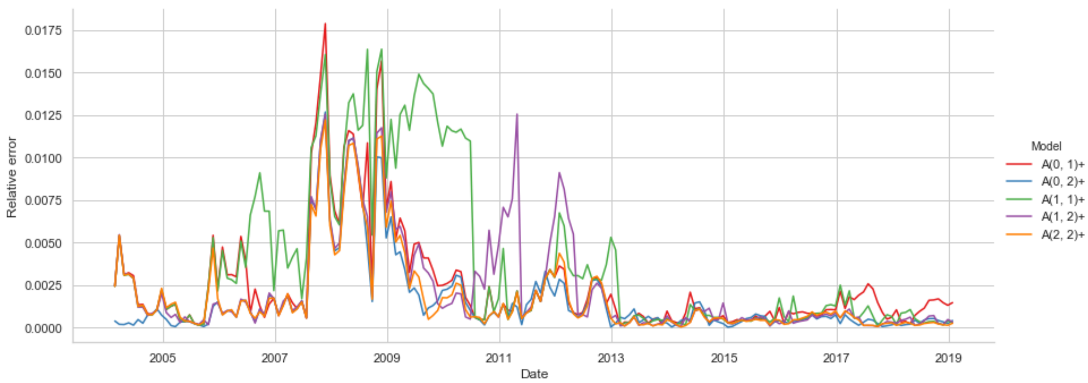


Figure 6.12: Time series of mean absolute relative errors in Euribor fitting.

Overall, two-factor models had the best performance and $A(1,1)+$ was significantly the worst in performance. The correlation parameter of model $A(0,2)+$ hows that the factors are heavily correlated. First the parameters have near perfect negative correlation but when the short rates get close to zero, the correlation switched to nearly perfect positive correlation. Since the $A(0,2)+$-model has a definitive out-performance over the other Gaussian $A(0,1)+$-model, this high correlation does not seem to imply that only a one factor is sufficient.

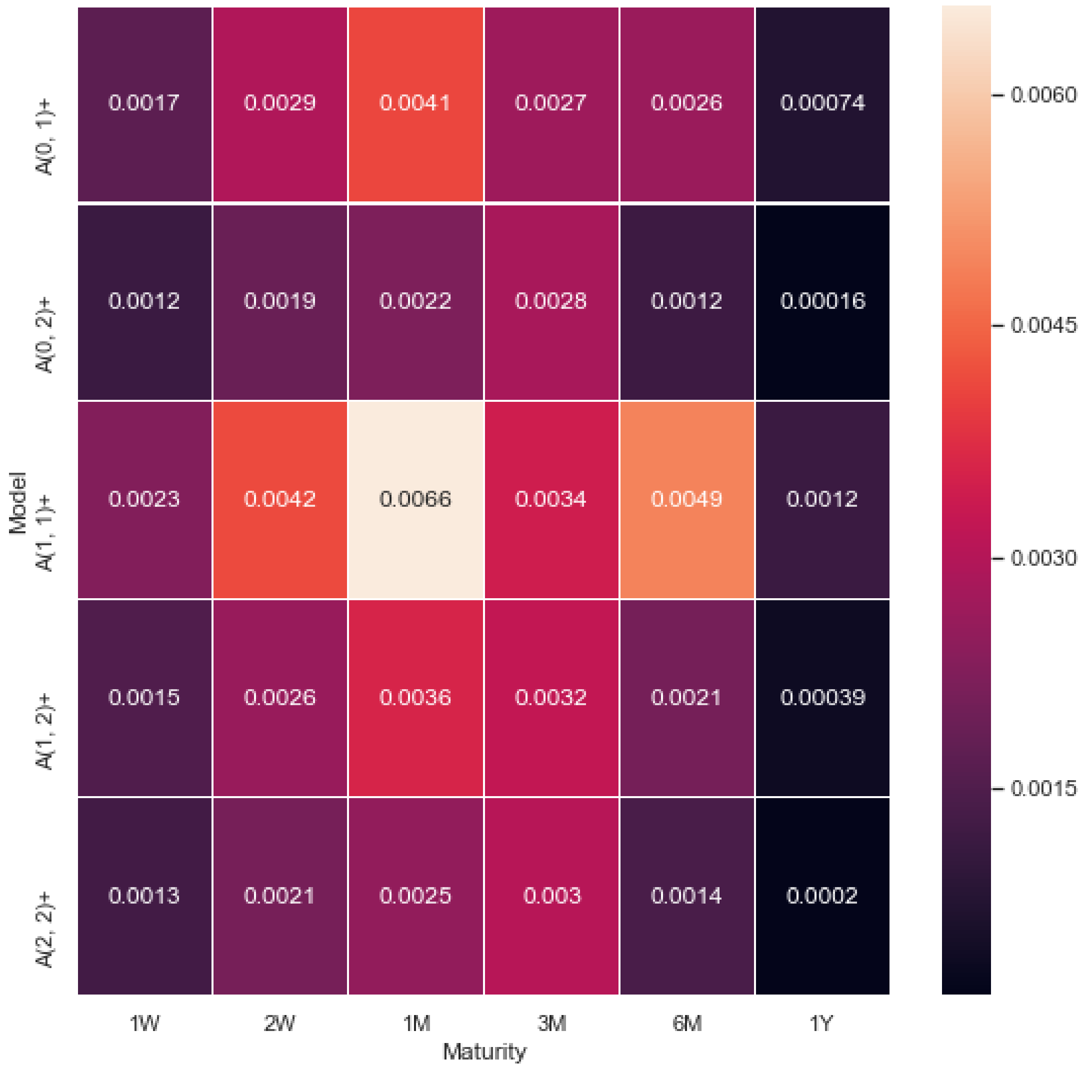


Figure 6.13: Errors in Euribor fitting by model and maturity.

Model	Mean absolute relative error (in percentages)
$A(0,1)+$	0.002464
$A(0,2)+$	0.001577
$A(1,1)+$	0.003749
$A(1,2)+$	0.002225
$A(2,2)+$	0.001755

As seen in Figure 6.12, time variability of error levels is high. In order to quantify thism the data was divided into four consecutive periods:

- Pre-crisis: from February 26, 2004 to September 26, 2008
- Surging rates during crisis: from September 26, 2008 to January 26, 2009

- Post-crisis rising rates: from January 26, 2009 to July 26, 2011
- Toward negative rates: from July 26, 2011 to January 26, 2019

The actual dates were chosen so that they occur during certain local minimum or maximum rates. We can clearly notice how badly all the models handle shorter and middle range maturities during the when the rates dropped rapidly, but two-factor models seem to get the one -year rat evolution quite right even if it has the largest absolute moves. During rising regimes all the models have problem. When the rates are negative an curve is very flat, all the models seem to get work fairly well.

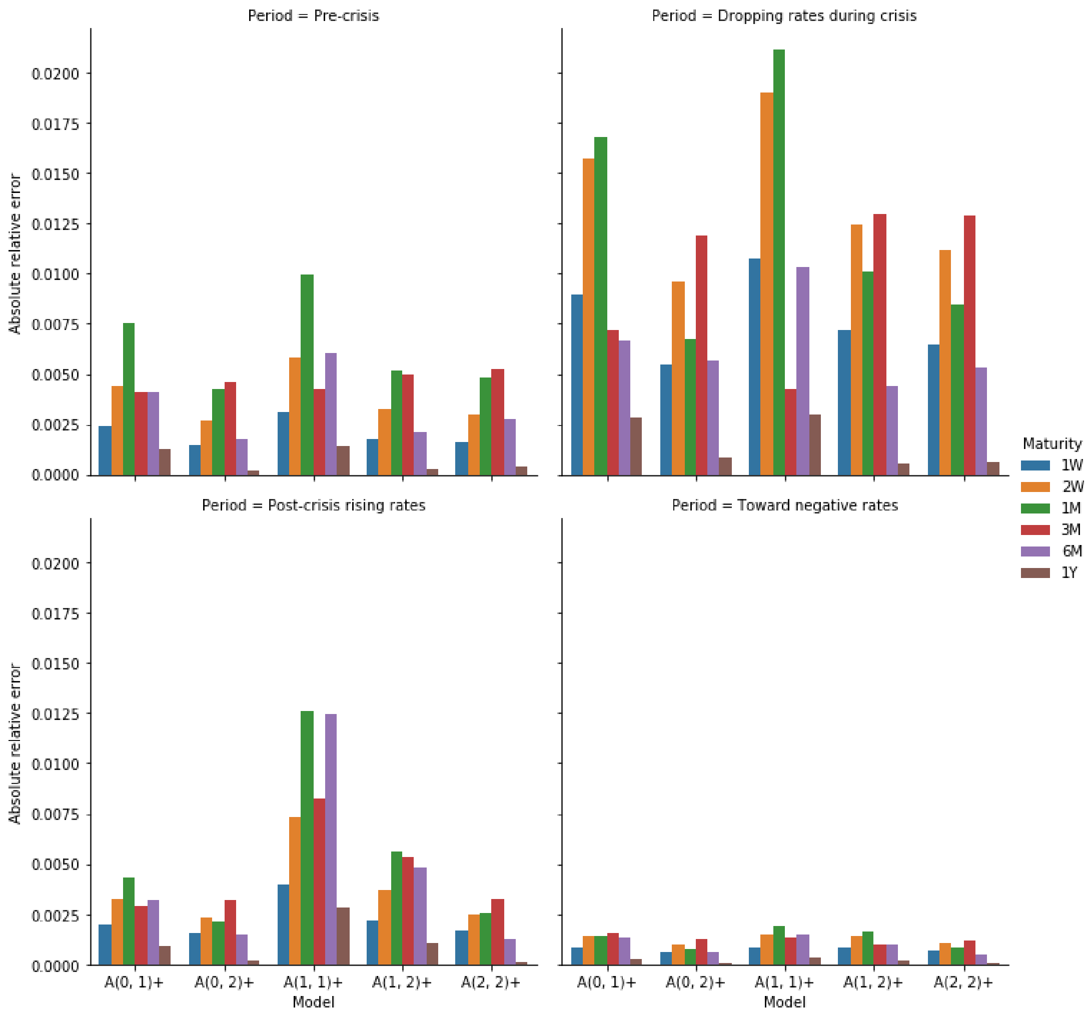


Figure 6.14: Errors in Euribor fitting by time period.

7. SUMMARY

The purpose of this thesis was to give an overview of arbitrage-free pricing methodology and affine short-rate processes used in interest rate modeling and credit risk. These kind of models have a long history starting from the late 1970's. Their heyday was pre-financial crisis of 2007-08. After the crisis, many significant changes have occurred in interest rate markets, for example:

- IBOR rates are not anymore considered as riskless and relevant spreads have widened.
- Credit adjustments are required for unsecured positions.
- Multi-curve pricing has been the industry standard at least for linear products such as swaps.
- Extraordinary monetary policies of central banks have caused negative interest rates, which were earlier considered impossible. As such, the possibility of negative rates in Vašíček–model was considered a flaw earlier.

Although short-rate models have been surpassed by market models in practice, it is an interest question consider how short-rate models fare in the current market structures.

Short-rate models have several theoretical short-comings. The first is conceptual, there is no actual instantaneous short-rate. It is purely theoretical concept created to explain how the interest rates are formed. Although having an unobservable process as a main ingredient of a theory is troubling, it can be forgiven if implied theoretical structure is otherwise logically sound and it can produce accurate results. The viability of Black-Scholes option pricing methodology is based on the assumption that the future volatility of the stock process can be accurately inferred, even if it is not actually observable. Same can be true for short-rate modeling of interest rates. If the observed term structure is coherent with the implications of a hypothesized short-rate process, then the model might be useful even if the model might be fundamentally wrong.

The second main theoretical short-coming of short-rate modeling is the fact that it is mainly concerned of a single point, the next infinitesimal future time-step. As such, it is not fat-fetched to hypothesize that short-rate models have hard time to explain complex temporal evolution of interest rate curve. For example, we have demonstrated that for affine short-rate models with one factor, the long-rates are perfectly correlated. This is a severe limitation but it can be mitigated with the introduction of multiple factors, time-varying parameters or technique of dynamic extension.

However, short-rate models are not without merits. They are conceptually simple to understand. Affine models are analytically tractable with explicit analytical bond pricing formulas. Some models even have explicit analytical bond option pricing formulas which can be converted to the price caps and floors.

The calibration of models to the market data was inconclusive in the sense that we did not achieve consistent results. For simple models without credit risk, although there were some precise calibrations, no model was consistently accurate. Since we employed stochastic optimization algorithms for calibration, there is no certainty that global minimums were found. The curse of dimensionality makes the optimization problem very hard for multi-factor models. Thus a bad fit does not indicate that the model is unsuitable for the observed data. Since single-factor models have a manageable number of parameters, we could expect optimization to be fairly dependable. For $A(0,1)+$–model, the alternative calibration replicated the original parameter values almost perfectly. For $A(0,1)+$–model, the alternative calibration produced significantly different parameter values but the accuracy was very similar. Therefore we can infer with reasonable confidence that $A(0,1)+$ and $A(1,1)+$–models can not necessarily fit post-crisis term-structures. Since some of the considered multi-factor models of family $A(M,N)+$ offered decent accuracy, we believe that two or three factor models could be used to fit the recent observed data. As the alternative calibrations led to significantly different parameters and calibration errors, the inference about model quality of multi-factor models of type $A(M,N)+$ is not reliable.

Since the calibration data included maturities ranging from overnight rate to 30-year rate, the observed rate structure is complex. Calibration to subset of these maturity ranges will likely produce significantly better fits. This reasoning is supported results of the dynamic calibration of Euribor rates ranging from one week to annual maturity, which generally give significantly better accuracy.

For simple models with credit risk, the calibration errors were large. However, calibration was only attempted with recent data and only models with two-factors were considered. As we had seen, single-factor models did not perform well in this environment for interest rate curves and these factors had to explain both the interest rate and spread curves. It is probably that models with more factors could work better.

The calibration methodology employed had severe short-comings. Although differential evolution has the desired ability to explore the optimization space, it probably wasted lots of function calls to explore infeasible regions. On the other, since the alternative optimizations tended not to converge original points, it seems that meta-parameters for the optimization were misspecified. Also it is not clear if differential evolution is the best choice for this kind of calibration. Particle swamp optimization or simulated annealing might have been better alternatives. The large number of solutions near the optimization borders suggest that those borders might have been inappropriate. On the other hand. Since the shifts were stopped at borders, the borders were also likely to be hit during optimization tries.

It could be interesting to test how well these affine models and their dynamic extensions compare against more modern models such as SABR when they are calibrated to recent volatility structures. Since descendants of LIBOR models are based on the assumption of log-normal distribution, negative interest rates causes problems that require unnatural solutions such as shifts or normality assumptions that may lower model quality. Negative rates are possible for affine models with Gaussian factors, but this probably does not compensate for inferior volatility fabrics of these models.

APPENDIX A. MATHEMATICAL APPENDIX

In this chapter, we shall review the intuition behind basic results in probability theory and stochastic calculus. We shall omit most of the measure theory needed and be pretty vague of the technical arguments.

In the following we let $(\Omega, \mathscr{F})$ be a probability space with a measure $\mathbb{P}$ be a probability measure. Here Ω is the state space, $\mathscr{F}$ is the σ-algebra of Ω. Let $\mathbb{Q}$ be also a probability measure on the space $(\Omega, \mathscr{F})$. If for all $A \in \mathscr{F}$, it holds that $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$, then we say that $\mathbb{Q}$ is absolutely continous with respect to $\mathbb{P}$ on $\mathscr{F}$ and we write $\mathbb{Q} \ll \mathbb{P}$. If $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$, then the measures are said to be equivalent on $\mathscr{F}$. Thus measures are equivalent if and only if their null sets are the same.

A.1. Characteristic function and Fourier transformation

A probability distribution function of random variable X is any measurable function f_X that satisfies

$$\mathbb{P}(X \in A) = \int_A f_X \, \mathrm{d}\mu \tag{A.1.1}$$

for all $A \in \mathscr{F}$, where μ is the Lebesgue measure. The characteristic function g_X is the function

$$g_X(\omega) = \mathrm{E}_{\mathbb{P}}\left(\mathrm{e}^{i\omega X}\right) \tag{A.1.2}$$

$$= \int_{-\infty}^{\infty} \mathrm{e}^{i\omega x} f_X(x) \, \mathrm{d}x \tag{A.1.3}$$

if the expectation exists. This is just the Fourier transformation of the probability distribution function. If the characteristic function g_X is integrable, then the inverse Fourier transformation gives

$$f_X = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-i\omega x} g_X x \, \mathrm{d}x. \tag{A.1.4}$$

According to Carr et al. 1999, Gil-Pelaez' Inversion gives that

$$\mathbb{P}(x < X) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}\left(\mathrm{e}^{-i\omega x} g_X(\omega)\right)}{\omega}\, \mathrm{d}\omega \tag{A.1.5}$$

and

$$\mathbb{P}(x \geq X) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{\mathrm{e}^{-i\omega x} g_X(\omega)}{i\omega}\right) \mathrm{d}\omega. \tag{A.1.6}$$

A.2. Radon-Nikodým-theorem and the change of measure

One of the basic tool of the probability measures is the Radon-Nikodým-theorem. For the proof, see any basic text on the measure theory (for example, Billingsley (1992, pp. 449-450)). We recall that a function f is $\mathscr{F}$-measurable if $\{\omega \in \Omega \mid f(\omega) \leq x\} \in \mathscr{F}$ for any $x \in \mathbb{R}$.

Theorem A.2.1 (Radon-Nikodým-theorem)**.** *If $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on measurable space $(\Omega, \mathscr{F})$ and $\mathbb{Q} \ll \mathbb{P}$, then there exists a non-negative $\mathscr{F}$-measurable function ξ such that*

$$\int_\Omega \xi \, \mathrm{d}\mathbb{P} < \infty \textit{ and} \tag{A.2.1}$$

$$\mathbb{Q}(A) = \int_A \xi \, \mathrm{d}\mathbb{P} \tag{A.2.2}$$

for all $A \in \mathscr{F}$. The function ξ is $\mathbb{P}$-unique and it is called as the Radon-Nikodým-derivate of $\mathbb{P}$ with respect to measure $\mathbb{Q}$ and filtration $\mathscr{F}$.

The Radon-Nikodým-derivate of $\mathbb{P}$ with respect to $\mathbb{Q}$ is denoted by

$$\xi = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \tag{A.2.3}$$

and alternatively we may write

$$\mathrm{d}\mathbb{Q} = \xi \mathrm{d}\mathbb{P}. \tag{A.2.4}$$

Since $\mathbb{Q}$ is a probability measure, it is clear that $\mathrm{E}_{\mathbb{P}}(\xi) = 1$.

If $X = 1_A$ for some $A \in \mathscr{F}$, then

$$\mathrm{E}_{\mathbb{Q}}(X) = \mathbb{Q}(A) = \int_A \xi \mathrm{d}\mathbb{P} = \mathrm{E}_{\mathbb{P}}(1_A \xi) = \mathrm{E}_{\mathbb{P}}(\xi X). \tag{A.2.5}$$

If X is integrable and $\mathscr{F}$-measurable random variable, then we may approximate it with simple functions and we may conclude the following important consequence of the Radon-Nikodým-derivate.

Lemma A.2.2. *Suppose that the function ξ is the Radon-Nikodým-derivate of $\mathbb{P}$ with respect to $\mathbb{Q}$. If X is integrable and $\mathscr{F}$-measurable random variable, then*

$$\mathrm{E}_{\mathbb{Q}}(X) = \mathrm{E}_{\mathbb{P}}(\xi X). \tag{A.2.6}$$

Conversely, if ξ is $\mathscr{F}$-measurable, integrable and non-negative function with $\mathrm{E}_{\mathbb{P}}(\xi) = 1$, then we may define a function $\mathbb{Q} : \mathscr{F} \to [0,1]$ by

$$\mathbb{Q}(A) = \int_A \xi \mathrm{d}\mathbb{P} = \mathrm{E}_{\mathbb{P}}(1_A \xi) \tag{A.2.7}$$

for all $A \in \mathscr{F}$. It is easy to see that $\mathbb{Q}$ is a probability measure on measurable space $(\Omega, \mathscr{F})$ and $\mathbb{Q} \ll \mathbb{P}$. Also

$$\xi = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}. \tag{A.2.8}$$

By Equation A.2.7,

$$\mathrm{E}_{\mathbb{Q}}(X) = \mathrm{E}_{\mathbb{P}}(X \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}) \tag{A.2.9}$$

holds for simple functions and, by limit argumentation, it holds for any integrable random variable. Heuristically

$$\int_A X \ \mathrm{d}\mathbb{Q} = \int_A X \ \mathrm{d}\mathbb{P} \ \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \tag{A.2.10}$$

$$= \int_A X \ \mathrm{d}\mathbb{P} \tag{A.2.11}$$

for all $A \in \mathscr{F}$.

A.3. Conditional expectation

The associated σ-algebra $\mathscr{F}$ can be seen as the known information structure. Random variables are $\mathscr{F}$-measurable functions which means that the sets[1] $\{ \ \omega \mid X(\omega) \leq a \ \} \in \mathscr{F}$ for all $a \in \mathbb{R}$. This may be interpreted as that $\mathscr{F}$-measurable functions are those functions whose outcome is known based on the information $\mathscr{F}$.

If $\mathscr{F} = \{\emptyset, \Omega\}$, then only constant functions are $\mathscr{F}$-measurable and knowning the value of random variable gives no information about the true state of the system ω. If

[1] These sets generate the Borel algebra of the reals.

$\emptyset \neq A \subset \Omega$ and $A \in \mathscr{F}$, then 1_A is $\mathscr{F}$-measurable function. So if we know the value of 1_A, we may deduce either $\omega \in A$ or $\omega \notin A$ although we may not have exact information about the true state ω of the random system Ω. So if $\mathscr{G}$ is a σ-sub-algebra of $\mathscr{F}$, then $\mathscr{F}$ carries more information than $\mathscr{G}$.

Let X be a $\mathscr{F}$-measurable random variable and $\mathscr{G}$ a σ-sub-algebra of $\mathscr{F}$ generated by partion $B_1, B_2, \dots, B_m$ of Ω. For simplicity, we assume that X is simple, meaning that

$$X = \sum_{i=1}^{n} x_i 1_{A_i} \tag{A.3.1}$$

where $x_i \in \mathbb{R}$ and $A_i \in \mathscr{F}$ for $i = 1, 2, \dots, n$ and the collection $A_1, A_2, \dots A_n$ is a partition of Ω. We denote $C_{ij} = A_i \cap B_j$.

If we know that the event B_j is true, then $X = x_i$ only if $C_{ij} \neq \emptyset$ and, in a sense, the average value of X will be

$$y_j = \sum_{i=1}^{n} x_i \frac{\mathbb{P}(C_{ij})}{\mathbb{P}(B_j)} \tag{A.3.2}$$

assuming that $\mathbb{P}(B_j) \neq 0$. If $\mathbb{P}(B_j) = 0$, then

$$\mathrm{E}(X 1_{B_j}) = 0 \tag{A.3.3}$$

and we set $y_j = 0$. Now we may define a new random variable

$$Y = \sum_{j=1}^{m} y_j 1_{B_j}. \tag{A.3.4}$$

Now Y is $\mathscr{G}$-measurable and integrable. Furthermore,

$$\begin{aligned}
\mathrm{E}(X 1_{B_j}) &= \sum_{i=1}^{n} \mathrm{E}(x_i 1_{A_i} 1_{B_j}) && \text{(A.3.5)}\\
&= \sum_{i=1}^{n} x_i \mathbb{P}(C_{ij}) && \text{(A.3.6)}\\
&= y_j \mathbb{P}(B_j) && \text{(A.3.7)}\\
&= \mathrm{E}(Y 1_{B_j}) && \text{(A.3.8)}
\end{aligned}$$

for all $j = 1, 2, \dots, m$. This motivates us to define the conditional expectation given a σ-sub-algebra $\mathscr{G}$.

For a fixed $\mathscr{F}$-measurable and integrable random variable X, the conditional expectation of X given $\mathscr{G}$ is the random variable $\mathrm{E}(X|\mathscr{G})$ with the following properties:

i) $\mathrm{E}(X|\mathscr{G})$ is $\mathscr{G}$-measurable and integrable,

ii) for every $G \in \mathscr{G}$,

$$\int_G X \mathrm{d}\mathbb{P} = \int_G \mathrm{E}(X|\mathscr{G}) \mathrm{d}\mathbb{P}. \tag{A.3.9}$$

Thus conditional expectation is a random variable and

$$\mathrm{E}(X) = \mathrm{E}(\mathrm{E}(X|\mathscr{G})). \tag{A.3.10}$$

We may use Radon-Nikodým-theorem or orthogonal projections in $\mathscr{L}^2$-space to prove the existance of a conditional expectations and it is unique $\mathbb{P}$-surely. In the following, we shall not always make the distinction between sets or random variables that match everywhere or just $\mathbb{P}$-everywhere.

Let X_i be the throw of a fair coin at the time i. So $X_i(\omega) \in \{0,1\}$ with equal probabilities. For simplicity, we consider only two time periods $i = 1,2$ and we code

$$\Omega = \{\ X_i X_j \mid i,j \in \{0,1\}\ \} = \{\ 00, 01, 10, 11\ \}, \tag{A.3.11}$$

with $\mathbb{P}(\omega) = 1/4$ for all $\omega \in \Omega$. Let $\mathscr{F} = \{\emptyset, \Omega\}$ and $X(ij) = X_1(i) + X_2(j)$. Now

$$\mathrm{E}_{\mathbb{P}}(X \mid \mathscr{F}) = 1 \tag{A.3.12}$$

since

$$\int_\emptyset X\ \mathrm{d}\mathbb{P} = 0, \tag{A.3.13}$$

$$\int_\Omega X\ \mathrm{d}\mathbb{P} = 1 \tag{A.3.14}$$

$$\tag{A.3.15}$$

If we know the result of the first throw, then we may pick

$$\mathscr{G} = \mathscr{F} \cup \{00, 01\} \cup \{10, 11\} \tag{A.3.16}$$

Now

$$\int_{\{00,01\}} X \,\mathrm{d}\mathbb{P} = \frac{X_1(0)+X_1(0)+X_2(0)+X_2(1)}{4} = \frac{1}{4}, \tag{A.3.17}$$

$$\int_{\{10,11\}} X \,\mathrm{d}\mathbb{P} = \frac{X_1(1)+X_1(1)+X_2(0)+X_2(1)}{4} = \frac{3}{4}, \tag{A.3.18}$$

$$\int_{\{00,01\}} \alpha \,\mathrm{d}\mathbb{P} = \frac{\alpha}{2}, \tag{A.3.19}$$

$$\int_{\{10,11\}} \beta \,\mathrm{d}\mathbb{P} = \frac{\beta}{2}, \tag{A.3.20}$$

$$\tag{A.3.21}$$

implies that Y defined by $Y(00) = Y(01) = \frac{1}{2}$ and $Y(10) = Y(11) = \frac{3}{2}$ is the conditional expectation of X over $\mathscr{G}$ as now

$$\int_G Y \,\mathrm{d}\mathbb{P} = \int_G X \,\mathrm{d}\mathbb{P} \tag{A.3.22}$$

for all $G \in \mathscr{G}$.

If $\mathscr{G} = \{\emptyset, \Omega\}$, then $\mathscr{G}$-measurable functions are constant functions. The integral over the empty set is zero for all integrable randon variables and

$$\int_\Omega X \mathrm{d}\mathbb{P} = \mathrm{E}(X) = \int_\Omega \mathrm{E}(X)\mathrm{d}\mathbb{P}. \tag{A.3.23}$$

We see that $\mathrm{E}(X|\{\emptyset,\Omega\}) = \mathrm{E}(X)$ and the conditional expectation gives no further information. If X is $\mathscr{G}$-measurable, then the equation A.3.9 is trivially satisfied and we see that $X = \mathrm{E}(X|\mathscr{G})$. In particular, $X = \mathrm{E}(X|\mathscr{F})$ as X is $\mathscr{F}$-measurable.

We present some of the basic properties of conditional expectations.

Theorem A.3.1. *Suppose that $\mathscr{G}$ and $\mathscr{H}$ are sub-σ-fields of $\mathscr{F}$ and X and Y are integrable random variables. Then*

i) $\mathrm{E}|\mathrm{E}(X|\mathscr{G})| \leq \mathrm{E}|X|$,

ii) $\mathrm{E}(aX + bY|\mathscr{G}) = a\mathrm{E}(X|\mathscr{G}) + b\mathrm{E}(Y|\mathscr{G})$ *for all* $a, b \in \mathbb{R}$,

iii) if $X \leq Y$, *then* $\mathrm{E}(X|\mathscr{G}) \leq \mathrm{E}(Y|\mathscr{G})$,

iv) $|\mathrm{E}(X|\mathscr{G})| \leq \mathrm{E}(|X|\,|\mathscr{G})$,

v) if $XY \in \mathscr{L}^1(\Omega, P)$ *and* Y *if* G*-measurable, then* $\mathrm{E}(XY|\mathscr{G}) = Y\,\mathrm{E}(X|\mathscr{G})$,

vi) if $\mathscr{H} \subseteq \mathscr{G}$, *then* $\mathrm{E}(\mathrm{E}(X|\mathscr{G})|\mathscr{H}) = \mathrm{E}(\mathrm{E}(X|\mathscr{H})|\mathscr{G}) = \mathrm{E}(X|\mathscr{H})$, *vii) if* $\sigma(X)$ *and* $\mathscr{G}$ *are independent, then* $\mathrm{E}(X|\mathscr{G}) = \mathrm{E}(X)$,

viii) if $P(G) \in \{0,1\}$ *for all* $G \in \mathscr{G}$*, then* $\mathrm{E}(X|\mathscr{G}) = \mathrm{E}(X)$,

For the proof, see Billingsley (1992, pp. 472–477).

A.4. Filtrations and martingales

A collection $(\mathscr{F}_t)$ of σ-sub-algebras of $\mathscr{F}$ is called a filtration if $\mathscr{F}_t \subseteq F_s$ for all $0 \leq t \leq s$. Informally a filtration presents the flow of information. We assume some standard technical conditions for the filtrations. Every $\mathbb{P}$-null set must be a member of $\mathscr{F}_0$ and

$$\mathscr{F}_t = \bigcap_{t<s} \mathscr{F}_s \tag{A.4.1}$$

for all $t \geq 0$.

A stochastic process X is a function $X : (\mathbb{R}_+ \cup \{0\}) \times \Omega \to \mathbb{R}$. It is often written as $X = (X(t))$, where the argument $\omega \in \Omega$ is dropped. It is then a collection of random variables with index set $\{t \geq 0\}$. We say that the process $(X(t))$ is adapted to the filtration $(\mathscr{F}_t)$ if $X(t)$ is $\mathscr{F}_t$-measurable for each t. Thus the variable $X(t)$ of an adapted process contains the information of the process that has been accumulated so far.

A $(\mathscr{F}_t)$-adapted stochastic process is a martingale if $\mathrm{E}_{\mathbb{P}}(|X(t)|) < \infty$ and

$$\mathrm{E}_{\mathbb{P}}(X(s) \mid \mathscr{F}_t) = X(t) \tag{A.4.2}$$

for all $0 \leq t < s < \infty$. Thus a martingale is a process, where the present value is the best estimate for the all future expected values given the past information contained in the process.

A.5. A stopping time and localization

A random variable $\tau : \Omega \to \mathbb{R}_+ \cup \{0\}$ is a stopping time with respect to the filtration $\mathscr{F}_t$ if

$$\{\ \tau \leq t\ \} = \{\omega \in \Omega \mid \tau(\omega) \leq t\} \in \mathscr{F}_t \tag{A.5.1}$$

holds for all $t \geq 0$. This is equivalent to the existence of $(\mathscr{F}_t)$-adapted random process $(X(t))$ that

$$X(t) = \begin{cases} 0, & t \leq \tau, \\ 1, & t > \tau. \end{cases} \tag{A.5.2}$$

If τ is time of a default, then the condition of the Equation A.5.1 means that at the time t the information in the filtration will tell if the default has occured or not, that is $\tau \leq t$ or not.

Stopping times are used to localize behavior. For example, a local martingale is a $(\mathscr{F}_t)$-adapted process if there is such a sequence (τ_n) of stopping times that

$$\mathbb{P}(\tau_n < \tau_{n+1}) = 1, \tag{A.5.3}$$

$$\mathbb{P}(\lim_{n\to\infty} \tau_n = \infty) = 1, \tag{A.5.4}$$

and the stopped process defined by

$$X_{\tau_n}(t) = X(\min(t, \tau_n)) \tag{A.5.5}$$

is a $(\mathscr{F}_t)$-martingale for all $n \geq 1$.

A.6. Brownian motion

This section is adopted from Billingsley (1992, pp. 530–545).

In order to keep notation efficient, we denote in this section

$$\mathrm{E}_t(\cdot) = \mathrm{E}_t(\,\cdot \mid \mathscr{F}_t). \tag{A.6.1}$$

We also write $W(t, \omega) = W(t)$.

Let $(\mathscr{F}_t)$ be a filtration of the probability space. A Brownian motion (or a Wiener process) $W(t), t \geq 0$ with respect to filtration $(\mathscr{F}_t)$ is a stochastic process satisfying the following

1. $W(0) = 0$ almost surely,

2. $W(t)$ is $\mathscr{F}_t$-measurable for each $t \geq 0$,

3. $t \mapsto W(t)$ is $\mathbb{P}$-surely continuous,

4. for any finite set of times $0 \leq t_1 < t_2 < \ldots < t_n \leq T$, the random variables

$$W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_n) - W(t_{n-1}) \tag{A.6.2}$$

are independent,

5. $W(s) - W(t) \sim N(0, s-t)$ for all $0 < t < s$.

These imply that $W(t) = W(t) - W(0) \sim N(0,t)$ and

$$\mathrm{Var}(W(s) - W(t) = \mathrm{E}\left((W(s) - W(t)^2\right) = s - t \tag{A.6.3}$$

for all $0 < t < s$.

A Brownian motion $W(t)$ is indeed a martingale in respect to the natural filtration since $\mathrm{E}_t(W(s) - W(0) = 0$ and

$$\mathrm{E}_t(W(s)) = \mathrm{E}_t(W(s) - W(t) + W(t)) = \mathrm{E}_t(W(s) - W(t) + W(t) = W(t). \tag{A.6.4}$$

Similarly $W(s)^2 = (W(s) - W(t)^2 + 2W(s)W(t) - W(t)^2$ implies that

$$\begin{aligned}
\mathrm{E}_t(W(s)^2 - s) &= \mathrm{E}_t\left((W(s) - W(t)^2 + 2W(s)W(t) - W(t)^2\right) - s && \text{(A.6.5)}\\
&= \mathrm{Var}_t(W(s) - W(t)) + 2\,\mathrm{E}_t(W(s))W(t) - W(t)^2 - s && \text{(A.6.6)}\\
&= s - t + W(t)^2 - s && \text{(A.6.7)}\\
&= W(t)^2 - t && \text{(A.6.8)}
\end{aligned}$$

meaning that $W(t)^2 - t$ is also a martingale.

If $f, g : [0,T] \to \mathbb{R}$ are functions, then the covariation of f and g up to time T is

$$\langle f, g\rangle(T) = \lim_{|\Pi| \to 0} \sum_{i=0}^{n-1} (f(t_{i+1}) - f(t_i))(g(t_{i+1}) - g(t_i)), \tag{A.6.9}$$

where $\Pi = \{t_0, t_1, \ldots, t_n\}$, $0 = t_0 < t_1 < \ldots < t_n = T$ is a partition with mesh $|\Pi| = \max_i(t_{i+1} - t_i)$. The quadratic variation of a function f up to time T is

$$\langle f\rangle_T = \langle f, f\rangle(T) = \lim_{|\Pi| \to 0} \sum_{i=0}^{n-1} (f(t_{i+1}) - f(t_i))^2. \tag{A.6.10}$$

If the function f has continuous derivate, then we may use intermediate value theorem to conclude that

$$\sum_{i=0}^{n-1}(f(t_{i+1})-f(t_i))^2 = \sum_{i=0}^{n-1}(f'(s_i))^2(t_{i+1}-t_i)^2 \tag{A.6.11}$$

$$\leq |\Pi| \sum_{i=0}^{n-1}(f'(s_i))^2(t_{i+1}-t_i), \tag{A.6.12}$$

for some $t_i \leq s_i \leq t_{i+1}$ and where

$$\sum_{i=0}^{n-1}(f'(s_i))^2(t_{i+1}-t_i) \to \int_0^T (f'(t))^2 \mathrm{d}t < \infty \tag{A.6.13}$$

as $|\Pi| \to 0$. Here we also used the continuity of f' to keep the integral finite. This implies that $\langle f \rangle_T = 0$ for a smooth function f.

For random processes the quadratic variation is defined when the limit in probability exists for any sequence of partitions.

Theorem A.6.1. *If $W = (W(t))$ is a Brownian motion, then $\langle W \rangle_T = T$ for all $T \geq 0$.*

Proof. First $\mathrm{E}\left((W(t_{i+1})-W(t_i))^2\right) = t_{i+1}-t_i$. We recall the fact that for a random variable $X \sim N(0,\sigma^2)$ we have $\mathrm{Var}(X^2) = 2\sigma^4$. Thus

$$\mathrm{Var}\left((W(t_{i+1})-W(t_i))^2\right) = 2(t_{i+1}-t_i)^2. \tag{A.6.14}$$

These and the independence of increments implies that

$$\mathrm{E}\left(\sum_{i=0}^{n-1}(W(t_{i+1})-W(t_i))^2\right) = T \tag{A.6.15}$$

and

$$\mathrm{Var}\left(\sum_{i=0}^{n-1}(W(t_{i+1})-W(t_i))^2\right) = 2\sum_{i=0}^{n-1}(t_{i+1}-t_i)^2 \leq 2\,|\Pi|\,T. \tag{A.6.16}$$

This means that $\langle W \rangle_T$ $\mathscr{L}^2$-converges to T. □

The Brownian motion accumulates one unit of quadratic variation per unit of time. The previous result is often written informally as $\mathrm{d}W(t)\mathrm{d}W(t) = dt$. Similarly we may calculate the covariation $\langle W, t \rangle_T = 0$. It is enough to note that

$$\left|\sum_{i=0}^{n-1}(W(t_{i+1})-W(t_i))(t_{i+1}-t_i)\right| \leq T \max_i |W(t_{i+1})-W(t_i)| \tag{A.6.17}$$

and by continuity of the paths we may force $\max_i |W(t_{i+1}) - W(t_i)|$ converge to zero. This we will use informally as $\mathrm{d}W(t)\mathrm{d}t = 0$. Since $f(t) = t$ is a smooth function, we have that $\langle t,t\rangle(T) = 0$ and thus $\mathrm{d}t\mathrm{d}t = 0$. Hence

$$\mathrm{d}W(t)\ \mathrm{d}W(t) = dt, \tag{A.6.18}$$

$$\mathrm{d}W(t)\ \mathrm{d}t = 0, \tag{A.6.19}$$

$$\mathrm{d}t\ \mathrm{d}t = 0. \tag{A.6.20}$$

A.7. Itô-integral

This section is adopted from Øksendal (2003, pp. 21–55).

We would like to calculate the integral of a stochastic function h with respect to a Brownian motion W over the time interval $[0,T]$. Let $(\mathscr{F}_t)$ be the filtration induced by the Brownian motion W. Since h and W are random and W is $\mathbb{P}$-nowhere differentiable, we need to tread carefully. If a $(\mathscr{F}_t)$ adapted function $h(\omega,t)$ is constant on each subinterval $[t_i,t_{i+1}]$ given a partition $0 = t_0 < t_1 < \ldots < t_n = T$ of $[0,T]$, then h is called simple. For a simple function we may write

$$I(h)(\omega) = \sum_{i=0}^{n-1} h(\omega,t_i)(W(t_{i+1},\omega) - W(t_i,\omega)). \tag{A.7.1}$$

The key here is that the function $h(\omega,t_i)$ is at the earliest moment and therefore Itô-integral is not forward looking. Now $I(h)$ itself is a random variable. If h is structured enough, then we may approximate it with simple function and define Itô-integral as the limit of this process, if it is exists. Without going into details, we note that if $(\mathscr{F}_t)$-adapted process satisfies[2] that

$$\mathrm{E}\left(\int_0^T h^2(\omega,u)\mathrm{d}u\right) < \infty \tag{A.7.2}$$

then the limit

$$I(t,\omega) = \int_0^t h(\omega,u)\ \mathrm{d}W(u,\omega) \tag{A.7.3}$$

exists and is called the Itô-integral of h from 0 to t.

Under these common assumptions, the Itô-integral $I(t)$ satifies the following

1. $I(t)$ is $(\mathscr{F}_t)$-adapted,

2. $I(t)$ is continuous,

[2]Unless otherwise noted, we shall always assume that this condition will be satisfies.

3. $I(t)$ is a martingale and $\mathrm{E}(I(t) \mid \mathscr{F}_0) = 0$ for all $t \geq 0$ and

4. $I(t)$ satisfies the Itô-isometry

$$\mathrm{E}\left(I(h)^2\right) = \mathrm{E}\left(\int_0^T h^2(\omega, u)\mathrm{d}u\right) < \infty. \tag{A.7.4}$$

The quadratic variation of Itô-integral up to time t is

$$\int_0^t h^2 \mathrm{d}s. \tag{A.7.5}$$

Nothing restricts Itô-integral to be one dimensional. If $W(t)$ is d-dimensional Wiener process and $h(t, \omega) : \mathbb{R} \times \Omega \to \mathbb{R}^{p \times d}$, then we demand that

$$\mathrm{E}\left(\int_0^t |h(s, \omega)|^2 \ \mathrm{d}s\right) < \infty \tag{A.7.6}$$

where

$$|h(s, \omega)|^2 = \mathrm{tr}(h(s, \omega)^\top h(s, \omega)). \tag{A.7.7}$$

If we relax the condition above, then we may not guarantee that the Itô-integral will be martingale even if the integral exists.

A.8. Itô processes and Itô's lemma

This material in this section is adopted from Øksendal (2003, pp. 21–55).

An adapted process X is called an Itô-process if it can be written as

$$X(t) = X(0) + \int_0^t \mu(s)\mathrm{d}s + \int_0^t \sigma(s)\mathrm{d}W(s), \tag{A.8.1}$$

where $\mu(t, \omega), \sigma(t, \omega)$ are adapted and integrable processes. This is usually written in more informal differential notation as

$$\mathrm{d}X(t) = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t). \tag{A.8.2}$$

The function $\mu(t)$ is called as the drift and the function $\sigma(t)$ is called as the volatility. The quadratic variation of the Itô process is

$$\langle X \rangle_t = \int_0^t \sigma^2(u)\mathrm{d}u. \tag{A.8.3}$$

This is consistent with the heuristic calculation

$$dX(t)\ dX(t) = (\mu(t)dt)^2 + 2\mu(t)\sigma(t)dtdW(t) + (\sigma(t)dW(t)^2 \tag{A.8.4}$$
$$= \sigma^2(t)dt. \tag{A.8.5}$$

Theorem A.8.1 (Itô's lemma)**.** *If a function $g : \mathbb{R} \times [0,T] \to \mathbb{R}$ is in the class $C^{2,1}$ and $Y(t) = g(t, X(t))$, where $X(t)$ is an Itô-process, then $Y(t)$ is also an Itô-process with the presentation*

$$dY(t) = \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\mu(t) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\sigma^2(t)\right)dt + \frac{\partial g}{\partial t}\sigma(t)dW(t). \tag{A.8.6}$$

The formula itself is a short-hand for

$$Y_t = Y_0 + \int_0^t g_t(u)du + \int_0^t g_x(u)dX_u + \frac{1}{2}\int_0^t g_{xx}(u)d\langle X\rangle_u, \tag{A.8.7}$$

where, for example, g_x is the partial differential of g with respect to first variable.

If Itô-process

$$dX(t) = \mu(t)dt + \sigma(t)dW(t). \tag{A.8.8}$$

has $\mu : \mathbb{R} \times \Omega \to \mathbb{R}^d$, $\mu : \mathbb{R} \times \Omega \to \mathbb{R}^{p\times d}$ and $W(t)$ is d-dimensional Wiener process, then the equation in Itô's lemma can be written as

$$dY = g_t dt + g_x \mu dt + g_x \sigma dW + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} g_{x_i,x_j}(\sigma^\top \sigma)_{i,j}dt, \tag{A.8.9}$$

where $g_x = \left(g_{x_1} g_{x_2} \cdots g_{x_p}\right)$.

Itô's lemma can be used to calculate integrals. If $g(x,t) = x^2$ and $X_t = W_t$, then

$$dg(W_t, t) = dt + 2W(t)dW(t), \tag{A.8.10}$$

which is the short-hand for

$$W(t)^2 = t + 2\int_0^t W(s)dW(s). \tag{A.8.11}$$

This means that

$$\int_0^t W(s)dW(s) = \frac{1}{2}\left(W(t)^2 - t\right). \tag{A.8.12}$$

A.9. Geometric Brownian motion

A stochastic process X is a geometric Brownian motion if it satisfies the stochastic differential equation

$$\mathrm{d}X(t) = \mu X(t)\mathrm{d}t + \sigma X(t)\mathrm{d}W(t), \tag{A.9.1}$$

where μ and $\sigma > 0$ are constants and W is a Brownian motion. The constant μ is the drift and σ is the diffusion (or volatility). The solution of this SDE can be calculated by using Itô's lemma with the function $g(x,t) = \log(x)$. Now $\frac{\partial g}{\partial x} = 1/x$ and $\frac{\partial^2 g}{\partial x^2} = -1/x^2$. Thus

$$\mathrm{d}\log X(t) = \left(0 + \mu X(t)/X(t) - \frac{1}{2}\sigma^2 X^2(t)/X^2(t)\right)\mathrm{d}t + \sigma X(t)/X(t)\mathrm{d}W(t) \tag{A.9.2}$$

$$= \left(\mu - \frac{1}{2}\sigma^2\right)\mathrm{d}t + \sigma\mathrm{d}W(t) \tag{A.9.3}$$

and integrating from 0 to t gives us

$$\log(X(t)/X(0)) = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t). \tag{A.9.4}$$

Thus

$$X(t) = X(0)\mathrm{e}^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)}. \tag{A.9.5}$$

This means that for given $X(0)$, the variable $X(t)$ is log-normally distributed with parameters $(\mu - \frac{1}{2}\sigma^2)t$ and $\sqrt{t}\sigma$ normalized by $X(0)$. Now the mean of $X(t)$ is

$$X(0)\exp\left((\mu - \frac{1}{2}\sigma^2)t + \frac{1}{2}(\sqrt{t}\sigma)^2\right) = x(0)\exp(\mu t) \tag{A.9.6}$$

and the variance is

$$X(0)^2(\exp(\sigma^2 t) - 1)\exp(2(\mu - \frac{1}{2}\sigma^2)t + t\sigma^2) = X(0)^2(\exp(\sigma^2 t) - 1)\exp(2\mu t). \tag{A.9.7}$$

A.10. Girsanov's theorem

For reference, see Øksendal (2003, pp. 161–171).

As a converse of Radon-Nikodým-theorem, if L almost surely positive and $\mathrm{E}_{\mathbb{P}}(L) =$

1, then we may define a new measure

$$\mathbb{Q}(A) = \mathrm{E}_{\mathbb{P}}(L1_A) \tag{A.10.1}$$

for all $A \in \mathscr{F}$ and now

$$L = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \tag{A.10.2}$$

is the Radon-Nikodým-derivate with respect to $\mathbb{P}$ and $\mathbb{Q}$.

Theorem A.10.1 (Girsanov's theorem)**.** *Let $W(t)$ be a Brownian motion in the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with respect to filtration $(\mathscr{F}_t)$ and assume that the process $\kappa(t)$ is an $(\mathscr{F}_t)$-adapted process. We define*

$$L(t) = \exp\left(\int_0^t \kappa(s)\mathrm{d}W(s) - \frac{1}{2}\int_0^t \kappa^2(s)\mathrm{d}s\right) \tag{A.10.3}$$

$$= \exp\left(\int_0^t \kappa(s)\mathrm{d}W(s - \frac{1}{2}\langle\kappa(s)\rangle_t\right), \tag{A.10.4}$$

If we assume that $\mathrm{E}_{\mathbb{P}}(L_T) = 1$, then the process

$$W^*(t) = W(t) - \int_0^t \kappa(s)\mathrm{d}s. \tag{A.10.5}$$

is a Brownian motion under the equivalent probability measure $\mathbb{Q}$ defined by the Equation A.10.1.

The process κ is called as the Girsanov kernel of the probability transformation. If we assume that the Girsanov kernel satisfies the Novikov Condition

$$\mathrm{E}_{\mathbb{P}} \exp\left(\frac{1}{2}\int_0^T \kappa^2(s)\mathrm{d}s\right) < \infty, \tag{A.10.6}$$

then the condition $\mathrm{E}_{\mathbb{P}}(L(T)) = 1$ is satisfied.

The process $L(t)$ defined in the equation A.10.3 is the solution to the stochastic differential equation

$$\mathrm{d}L(t) = \kappa(t)L(t)\mathrm{d}W(t) \tag{A.10.7}$$

with the condition $L(0) = 1$.

The significant consequence of the theorem is the fact that the drift of stochastic process is very malleable. If

$$dX(t) = \mu(t)dt + \sigma(t)dW(t), \tag{A.10.8}$$

where $\sigma(t) > 0$, then we choose $\kappa(t) = -(\mu(t) - a(t))/\sigma(t)$. If the Girsanov kernel satisfies assumptions of the Girsanov's theorem, then we have a equivalent measure $\mathbb{Q}$ under which

$$W^*(t) = W(t) - \int_0^t \kappa(s)ds \tag{A.10.9}$$

is a Brownian motion. Thus $dW(t) = dW^*(t) + \kappa(t)dt$ and

$$dX(t) = \mu(t)dt + \sigma(t)(dW^*(t) + \kappa(t)dt) = a(t)dt + \sigma(t)dW^*(t) \tag{A.10.10}$$

under the measure $\mathbb{Q}$. We see that the measure change leaves diffusion unchanged, but the measure may be changed almost at the will. The Brownian motions $W(t)$ and $W^*(t)$ are not the same, but a priori their statistical properties are the same. Especially if $a(t) \equiv 0$, then the process $X(t)$ is driftless under the measure $\mathbb{Q}$. Thus

$$E_{\mathbb{Q}}(X(T)|\mathscr{F}_t) = X(t) \tag{A.10.11}$$

and it is a martingale.

A.11. Martingale representation theorem

For reference, see Wu (2009, pp.38–39)

Suppose that the filtration $(\mathscr{F}_t)$ is generated by a Brownian motion $W(t)$. One of the properties of the Itô-integral was that

$$I(t) = \int_0^t h(s)dW(s) \tag{A.11.1}$$

was a martingale, when $h(t)$ satisfies Equation A.7.2. Martingale representation states that if M is a square integrable martingale, that is

$$E_{\mathbb{P}}(M^2) < \infty, \tag{A.11.2}$$

then

$$M(t) = \int_0^t h(s)dW(s), \tag{A.11.3}$$

where $h(t)$ is $(\mathscr{F}_t)$-adapted and it satisfies Equation A.7.2. Thus

$$\mathrm{d}M(t) = h(t)\mathrm{d}W(t). \tag{A.11.4}$$

A.12. Feynman-Kac theorem

For reference, see Øksendal (2003, pp. 145–147).

The Feynman-Kac theorem states that the solution of the partial differential equation

$$\frac{\partial V(t,x)}{\partial t} + \frac{\partial V(t,x)}{\partial x} f(x) + \frac{1}{2}\frac{\partial^2 V(t,x)}{\partial x^2}\sigma^2(x) = rV(t,x) \tag{A.12.1}$$

with the terminal boundary condition $V(T,x) = g(x)$ is

$$V(t,x) = \mathrm{e}^{-r(T-t)}\,\mathrm{E}_{\mathbb{Q}}\left(g(X_T) \mid X(t) = x\right), \tag{A.12.2}$$

where the process $X(t) = x$ satisfies

$$\mathrm{d}X(s) = f(X(s))\mathrm{d}s + \sigma(X(s))\mathrm{d}W(s) \tag{A.12.3}$$

under the probability measure $\mathbb{Q}$, where $W(s)$ is a Brownian motion under the measure $\mathbb{Q}$.

A.13. Partial information

This section is based on McNeil, Frey, and Embrechts (2010, pp. 393–400).

We assume that the σ-algebra $(\mathscr{F}_t)$ presents partial market information without default and

$$\mathscr{H}_t = \sigma(1_{\{\zeta \le s\}} | s \le t) = \sigma(H(s) | s \le t) \tag{A.13.1}$$

is the knowledge of the default up to time t. By

$$\mathscr{F}_t \vee \mathscr{H}_t \tag{A.13.2}$$

we denote the smallest σ-algebra containing $\mathscr{F}_t$ and $\mathscr{H}_t$. Also

$$\mathscr{F}_\infty = \bigvee_t \mathscr{F}_t \tag{A.13.3}$$

the smallest σ-algebra containing all algebras $\mathscr{F}_t$.

Lemma A.13.1. *For every $A \in \mathscr{F}_t \vee \mathscr{H}_t$, there exists such $B \in \mathscr{F}_t$ that*

$$A \cap \{\zeta > t\} = B \cap \{\zeta > t\}, \tag{A.13.4}$$

where $t \in \mathbb{R}_+$.

Proof. We consider the filtration given by

$$\mathscr{G}_t = \{A \in \mathscr{F}_t \vee \mathscr{H}_t \mid A \cap \{\zeta > t\} = B \cap \{\zeta > t\} \text{ for some } B \in \mathscr{F}_t\} \tag{A.13.5}$$

and it is sufficient to show that $\mathscr{F}_t \vee \mathscr{H}_t \subseteq \mathscr{G}_t$. It is clear that $\mathscr{F}_t \subseteq \mathscr{G}_t$. If $A \in \mathscr{H}_t$, then $A \cap \{\zeta > t\}$ is either $\emptyset$ or $\{\zeta > t\}$ and it follows that $\mathscr{H}_t \subseteq \mathscr{G}_t$.

Trivially $\Omega \in \mathscr{G}_t$ and it is also easy to see that $\mathscr{G}_t$ is closed under countable unions. If $A \in \mathscr{G}_t$ and $B \in \mathscr{F}_t$ satisfies the equation (A.13.4), then

$$A^{\mathsf{c}} \cup \{\zeta \le t\} = B^{\mathsf{c}} \cup \{\zeta \le t\} \tag{A.13.6}$$

and

$$\begin{aligned} A^{\mathsf{c}} \cap \{\zeta > t\} &= (A^{\mathsf{c}} \cup \{\zeta \le t\}) \cap \{\zeta > t\} && \text{(A.13.7)} \\ &= (B^{\mathsf{c}} \cup \{\zeta \le t\}) \cap \{\zeta > t\} && \text{(A.13.8)} \\ &= B^{\mathsf{c}} \cap \{\zeta > t\} && \text{(A.13.9)} \end{aligned}$$

which implies that $A^{\mathsf{c}} \in \mathscr{G}_t$. Hence, $\mathscr{G}_t$ is a σ-algebra and $\mathscr{F}_t \vee \mathscr{H}_t \subseteq \mathscr{G}_t$. □

This can be used to show the following important result.

Lemma A.13.2. *If X is non-negative integrable random variable, then*

$$\mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta > t\}} X | \mathscr{F}_t \vee \mathscr{H}_t\right) = \frac{1_{\{\zeta > t\}}}{\mathbb{Q}(\zeta > t | \mathscr{F}_t)} \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta > t\}} X | \mathscr{F}_t\right). \tag{A.13.10}$$

Proof. This argument follows closely the result from McNeil et al. (2010, p. 396). Let $A \in \mathscr{F}_t \vee \mathscr{H}_t$. By Lemma A.13.1, there is $B \in \mathscr{F}_t$ such that

$$1_A 1_{\{\zeta > t\}} = 1_B 1_{\{\zeta > t\}}. \tag{A.13.11}$$

By this and the definition of conditional expectation we have

$$\int_A 1_{\{\zeta>t\}} X \mathbb{Q}(\zeta>t|\mathscr{F}_t)\,\mathrm{d}\mathbb{Q} = \int_B 1_{\{\zeta>t\}} X \mathbb{Q}(\zeta>t|\mathscr{F}_t)\,\mathrm{d}\mathbb{Q} \tag{A.13.12}$$

$$= \int_B \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} X|\mathscr{F}_t\right) \mathbb{Q}(\zeta>t|\mathscr{F}_t)\,\mathrm{d}\mathbb{Q} \tag{A.13.13}$$

$$= \int_B 1_{\{\zeta>t\}}\, \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} X|\mathscr{F}_t\right)\,\mathrm{d}\mathbb{Q} \tag{A.13.14}$$

$$= \int_A 1_{\{\zeta>t\}}\, \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} X|\mathscr{F}_t\right)\,\mathrm{d}\mathbb{Q}. \tag{A.13.15}$$

Since $A \in \mathscr{F}_t \vee \mathscr{H}_t$ is arbitrary, we have that

$$\mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} X \mathbb{Q}(\zeta>t|\mathscr{F}_t)|\mathscr{F}_t \vee \mathscr{H}_t\right) = 1_{\{\zeta>t\}}\, \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} X|\mathscr{F}_t\right), \tag{A.13.16}$$

where we have used facts $1_{\{\zeta>t\}} \in \mathscr{H}_t$ and $\mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} X|\mathscr{F}_t\right) \in \mathscr{F}_t$. As $\mathbb{Q}(\zeta>t|\mathscr{F}_t)$ is $\mathscr{F}_t \vee \mathscr{H}_t$-measurable, it can be taken out of the expectations. Since it is non-zero, we have the claim. □

Theorem A.13.3. *Let $T>t$. If X is non-negative integrable $\mathscr{F}_T$-measurable random variable, then*

$$\mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>T\}} X|\mathscr{F}_t \vee \mathscr{H}_t\right) = \frac{1_{\{\zeta>t\}}}{\mathbb{Q}(\zeta>t|\mathscr{F}_t)}\, \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>T\}} X|\mathscr{F}_t\right). \tag{A.13.17}$$

Proof. We will consider variable $Y = 1_{\{T>\zeta\}} X$. If $T>t$, then

$$1_{\{T>\zeta\}} X = Y = 1_{\{t>\zeta\}} Y, \tag{A.13.18}$$

and by the assumptions we may use Lemma A.13.2. Hence

$$\mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>T\}} X|\mathscr{F}_t \vee \mathscr{H}_t\right) = \mathrm{E}_{\mathbb{Q}}\left(1_{\{t>\zeta\}} Y|\mathscr{F}_t \vee \mathscr{H}_t\right) \tag{A.13.19}$$

$$= \frac{1_{\{\zeta>t\}}}{\mathbb{Q}(\zeta>t|\mathscr{F}_t)}\, \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} Y|\mathscr{F}_t\right) \tag{A.13.20}$$

$$= \frac{1_{\{\zeta>t\}}}{\mathbb{Q}(\zeta>t|\mathscr{F}_t)}\, \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>T\}} X|\mathscr{F}_t\right) \tag{A.13.21}$$

as required. □

A.14. Doubly stochastic default time

This section is based on the presentation from Filipović (2009, pp. 229 – 235).

The introduce the following technical assumptions:

(DS1) There exists a non-negative $\mathscr{F}_t$-progressive process $\lambda(t)$ such that

$$\mathbb{Q}(\zeta > t|\mathscr{F}_t) = \mathrm{e}^{-\int_0^t \lambda(s)\mathrm{d}s} \tag{A.14.1}$$

(DS2) For all $t \geq 0$, it holds that

$$\mathbb{Q}(\zeta > t|\mathscr{F}_t) = \mathbb{Q}(\zeta > t|\mathscr{F}_\infty), \tag{A.14.2}$$

where

$$\mathscr{F}_\infty = \bigvee_{t\geq 0} F_t. \tag{A.14.3}$$

The stopping times satisfying both (DS1) and (DS2) are doubly stochastic stopping times.

The condition (DS1) means that with only partial market information $\mathscr{F}_t$, the exact default time is never known as

$$0 < \mathbb{Q}(\zeta > t|\mathscr{F}_t) < 1 \tag{A.14.4}$$

for all $t \geq 0$. Hence $\mathscr{F}_t \neq \mathscr{F}_t \vee \mathscr{G}_t$ and ζ is not a stopping time under the filtration $(\mathscr{F}_t)$.

Now the Theorem A.13.3 directly implies the following.

Theorem A.14.1. *If we assume (DS1) and $T > t$, then*

$$\mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>T\}}X|\mathscr{F}_t \vee \mathscr{H}_t\right) = 1_{\{\zeta>t\}}\mathrm{e}^{\int_0^t \lambda(s)\mathrm{d}s}\left(1_{\{\zeta>T\}}X|\mathscr{F}_t\right) \tag{A.14.5}$$

holds for every non-negative integrable $\mathscr{F}_T$-measurable random variable X.

The previous Theorem is important, because we usually assume that the short rate $r(t)$ is adapted to partial information $(\mathscr{F}_t)$. In pricing instruments that are sensitive to credit risk, we have to start with full market information but the Theorem shows how we may switch back to partial information set.

Since

$$\begin{aligned}
\mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}}(1-1_{\{\zeta>T\}})|\mathscr{F}_t\right) &= \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} - 1_{\{\zeta>T\}}|\mathscr{F}_t\right) && \text{(A.14.6)}\\
&= \mathrm{E}_{\mathbb{Q}}\left(\mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} - 1_{\{\zeta>T\}}|\mathscr{F}_T\right)|\mathscr{F}_t\right) && \text{(A.14.7)}\\
&= \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} - \mathbb{Q}(\zeta > T|\mathscr{F}_T)|\mathscr{F}_t\right) && \text{(A.14.8)}
\end{aligned}$$

we know now that if (DS1) holds, then

$$\mathbb{Q}(t < \zeta \le T|\mathscr{F}_t \vee \mathscr{H}_t) = \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}} - 1_{\{\zeta>T\}}|\mathscr{F}_t \vee \mathscr{H}_t\right) \tag{A.14.9}$$

$$= \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}}(1 - 1_{\{\zeta>T\}})|\mathscr{F}_t \vee \mathscr{H}_t\right) \tag{A.14.10}$$

$$= \frac{1_{\{\zeta>t\}}}{\mathbb{Q}(\zeta > t|\mathscr{F}_t)} \mathrm{E}_{\mathbb{Q}}\left(1_{\{\zeta>t\}}(1 - 1_{\{\zeta>T\}})|\mathscr{F}_t\right) \tag{A.14.11}$$

$$= 1_{\{\zeta>t\}} \mathrm{E}_{\mathbb{Q}}\left(1 - \mathrm{e}^{-\int_t^T \lambda(s)\mathrm{d}s}|\mathscr{F}_t\right) \tag{A.14.12}$$

Thus we may approximate that

$$\mathbb{Q}(t < \zeta \le t + \mathrm{d}t|\mathscr{F}_t \vee \mathscr{H}_t) \approx 1_{\{\zeta>t\}}\lambda(t)\mathrm{d}t. \tag{A.14.13}$$

The condition (DS2) is equivalent to the condition that every $(\mathscr{F}_t)$-martingale is also a $(\mathscr{F}_t \vee \mathscr{H}_t)$-martingale. See, for example, Filipović 2009.

A.15. A gaussian calculation

For a continuous function g, we may approximate $\int_0^t g(s)\mathrm{d}s$ by the sequence

$$\frac{t}{n}\sum_{k=1}^{n} g\left(\frac{kt}{n}\right) \tag{A.15.1}$$

Theorem A.15.1. *If X is continuous and Gaussian stochastic process, then the stochastic process*

$$Y = \int_0^t X_s \mathrm{d}s \tag{A.15.2}$$

also has Gaussian distribution with

$$\mathrm{E}Y = \int_0^t \mathrm{E}X_s \mathrm{d}s, \tag{A.15.3}$$

$$\mathrm{Var}\,Y = \int_0^t \int_0^t \mathrm{Cov}(X_s, X_u)\mathrm{d}s\mathrm{d}u. \tag{A.15.4}$$

Proof. By using the idea above, we may approximate the distribution of Y by the distributions of

$$Y_n = \frac{t}{n}\sum_{k=1}^{n} X\left(\frac{kt}{n}\right), \tag{A.15.5}$$

which are gaussian as they are linear combinations of gaussian variables. Since the expectation is a linear operation, we have

$$\mathrm{E}Y_n = \frac{t}{n}\sum_{k=1}^{n} \mathrm{E}X\left(\frac{kt}{n}\right) \to \int_0^t \mathrm{E}X_s \mathrm{d}s \tag{A.15.6}$$

and

$$\mathrm{Var}\,Y_n = \frac{t}{n}\sum_{k=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}\left(X\left(\frac{kt}{n}\right), X\left(\frac{jt}{n}\right)\right) \to \int_0^t \int_0^t \mathrm{Cov}(X_s, X_u)\mathrm{d}s\mathrm{d}u. \tag{A.15.7}$$

□

A.16. Differential evolution

Differential evolution (DE) is a class of optimization methods based on evolutionary algorithms. It was introduced by Storn 1996 and Storn and Price 1997. DE requires no prior knowledge about the optimization problem. On the other hand, this means that after the algorithm has ran its course, we have no guarantees that the result is actually an optimal. It can be used to canvas large areas of the solution space. Therefore it can be useful to find the initial guess for other optimization algorithms that are sensitive to the precision of the initial value.

Suppose that $f : D \to \mathbb{R}$ is the fitness function that is to be minimized. Here D is a hypercube in real space $\mathbb{R}^n$.

The basic algorithm starts be selection of the initial population $I_0 \subset D$. Let I_k be the the population of k:th generation. The for every $x \in I_k$ we generate a distinct alternative candidate z has some heritage with x. If $f(z) < f(x)$, then we replace x with z in $(k+1)$:th generation. Therefore the fitness of the next generation as at least as good as the previous. The canon way of choosing the candidate is as follows:

1. For all $x \in I_k$ pick three distinct points a, b, c from $I_k \setminus \{x\}$.
2. Let $y = a + F(b - c)$ and randomly pick an index $j = 1, 2, \ldots, n$. We generate an evolutionary agent $z = (z_1, z_2, \ldots, z_n)$ by having $z_j = y_j$. For z_l, where $j \neq l \in \{1, 2, \ldots, n\}$ we pick y_l with the probability of C and otherwise x_l. Thus $z \neq x$. If $z \notin D$ then a new candidate is picked or it is scaled back to the search space.

The number F is called the differential weight and usually $F \in [0, 2]$. The probability C is the crossover probability. The choice of these parameters obviously influence the convergence. For example, small differential weights and crossover probabilities causes the population to converge quickly.

Since the basic setup of algorithm is very flexible, there are many variants. In the following, we outline the algorithm used in the data analysis of this thesis.

The initial population is chosen with uniform distribution. The size of the population is 1024 for models without default and 512 otherwise. For every 10 (or 5 for models with default) generations, a random check is made to see if there is a culling . The probability of this grows with each generations. The size of the culling is inversely proportional to the amount of replaces during the previous cycle. Only the members with worst fitness are removed. However, the population will always has at least 32 members.

For the candidate vectors we use two different strategies. Either we pick five distinct a, b, c, d, e and

$$z = a + F(b - c) + F(d - e) \tag{A.16.1}$$

or we pick four distinct candidates p, a, b, c and

$$z = a + F(p - a) + F(b - c) \tag{A.16.2}$$

where p is in the top 5% of the candidates. This first behaviour increases the chances of exploration while the latter approach will boost convergence. The algorithm is set up so that earlier exploration is preferred while later on the convergence is favored. Differential weight and the crossover probability are choosen randomly for each candidate. The algorithm is more likely to choose values that promote convergence later on the run.

APPENDIX B. CHARTS AND GRAPHS

B.1. Graphical presentation of the initial curve data

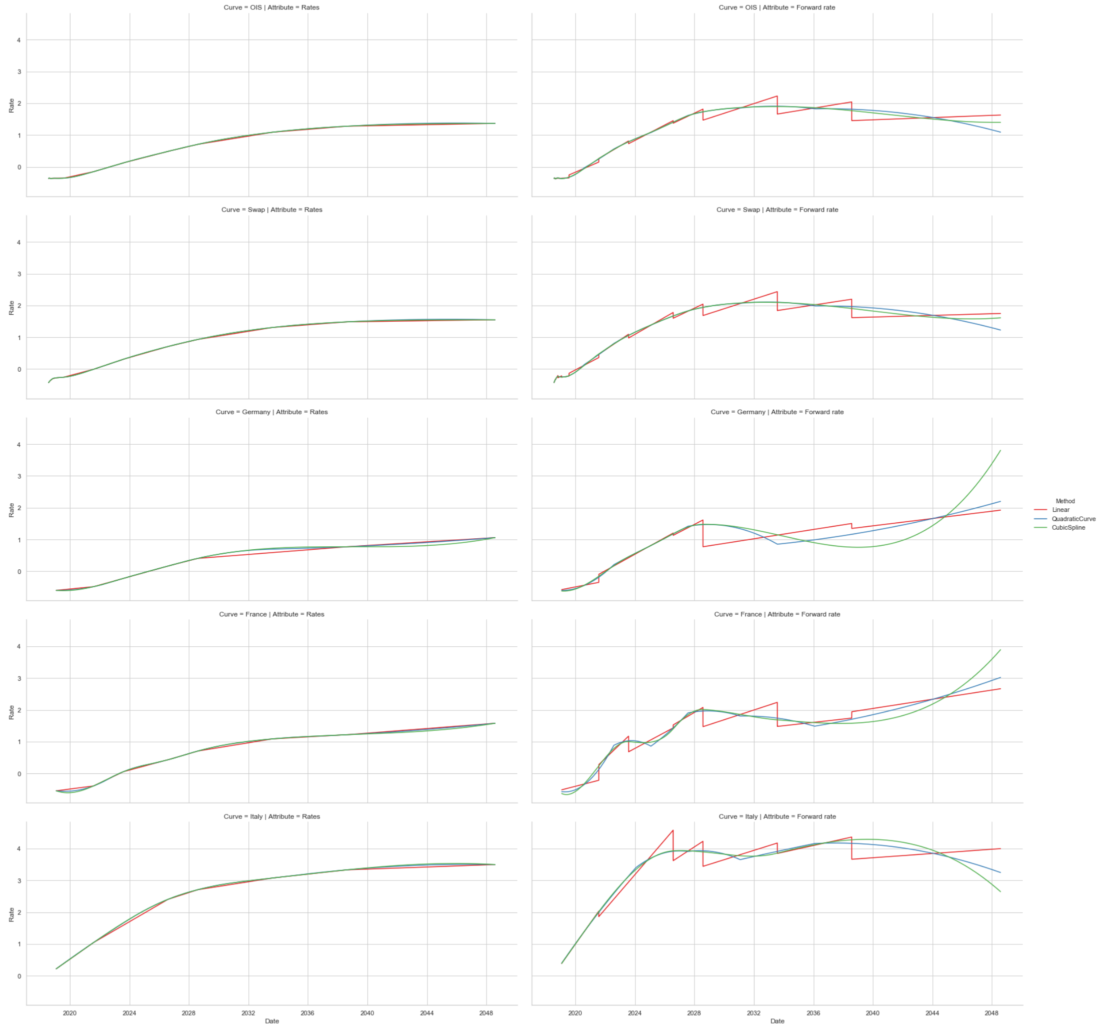


Figure B.1: Interpolated interest rates and instantaneous forward rate curves

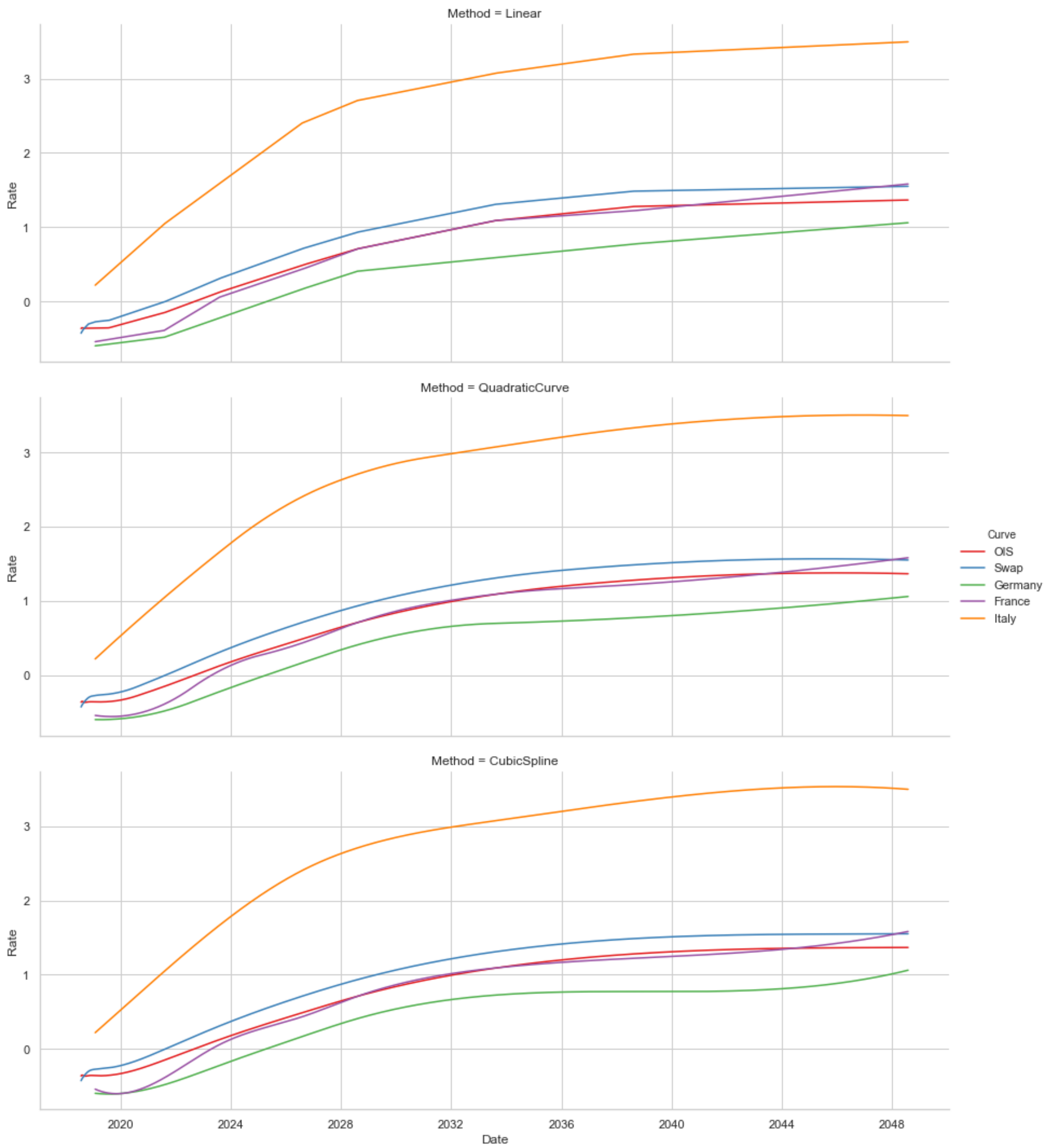


Figure B.2: Interpolated interest rates

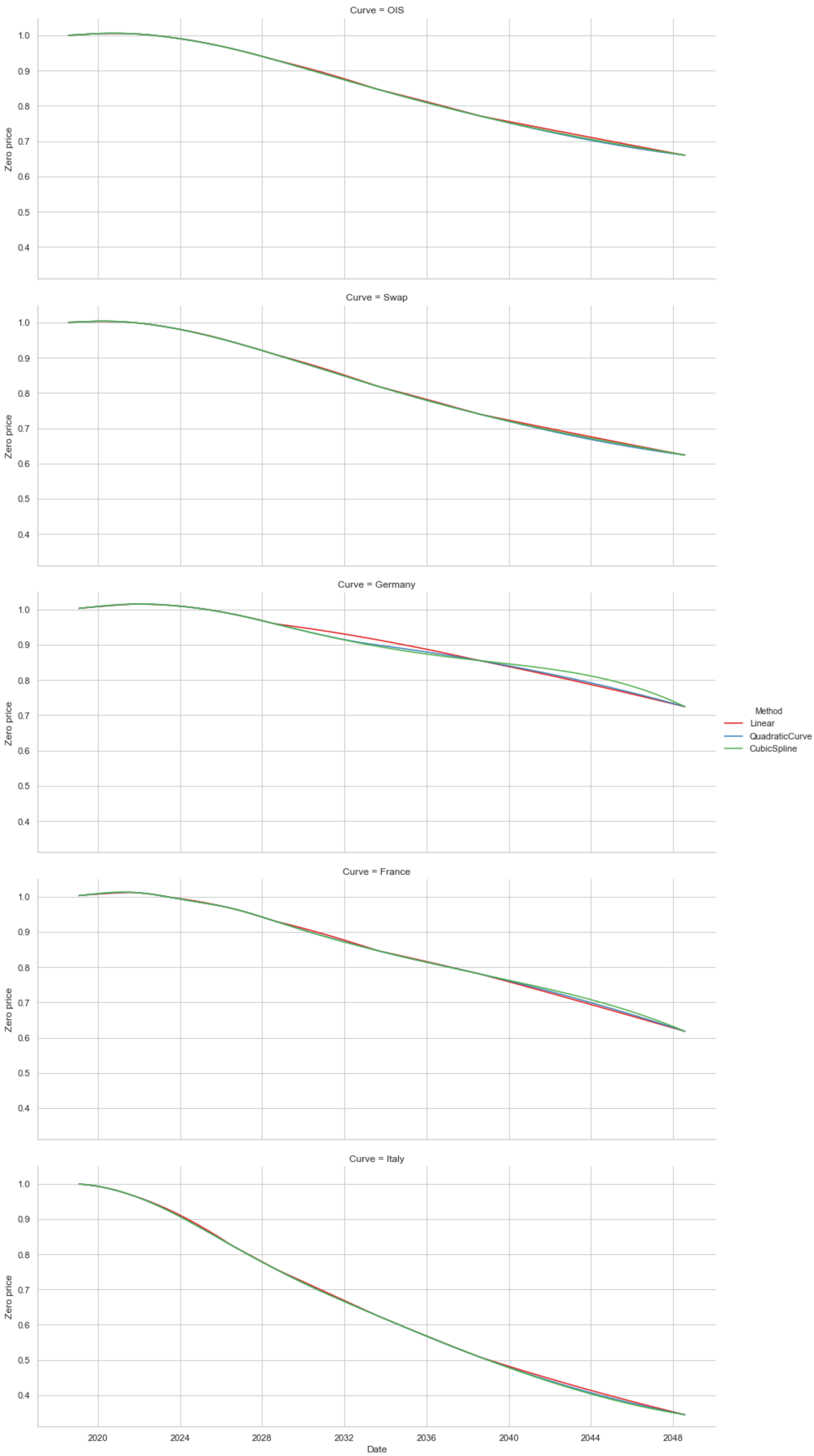


Figure B.3: Theoretical zero coupon bond prices inferred from the rate data by using

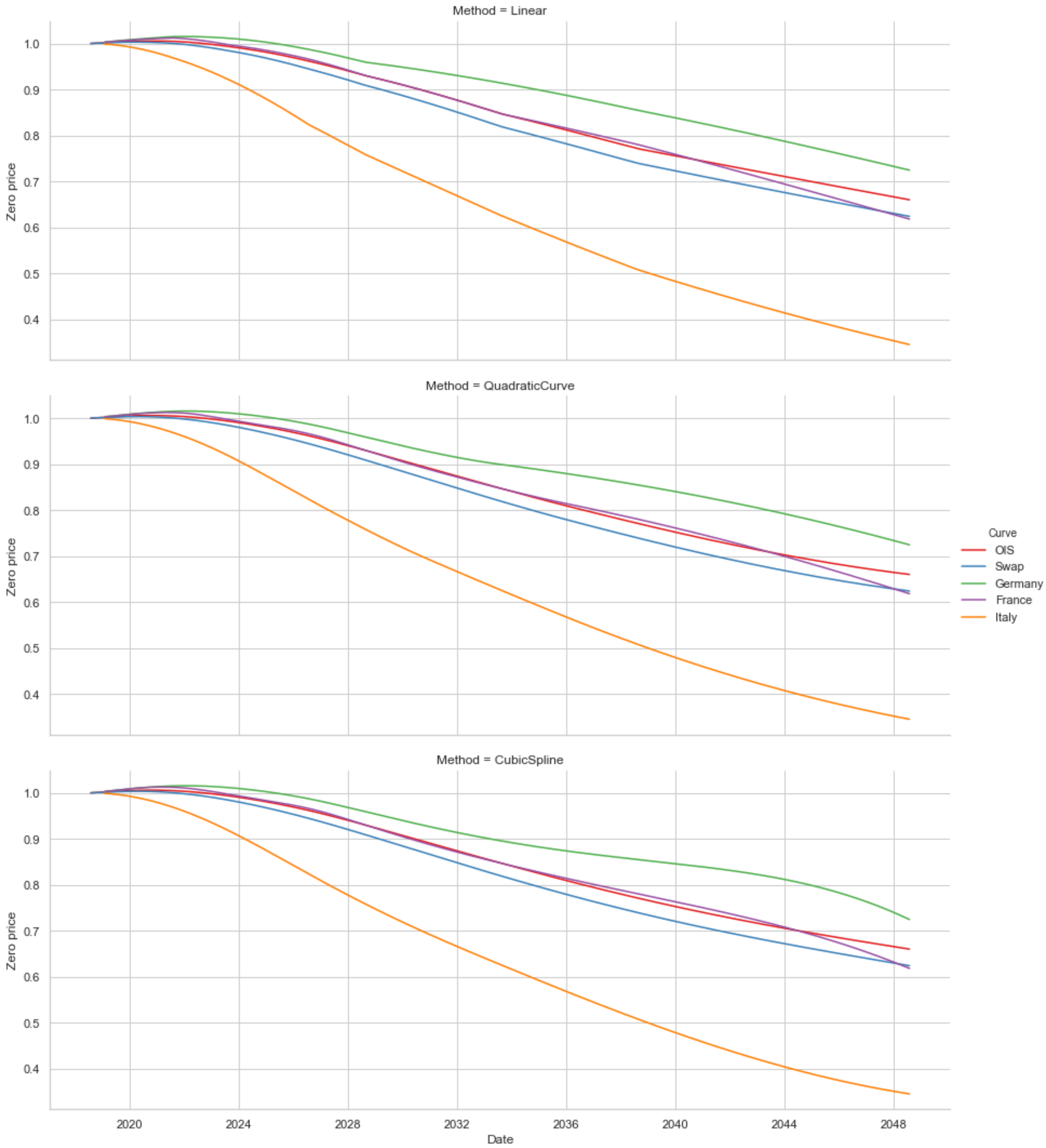
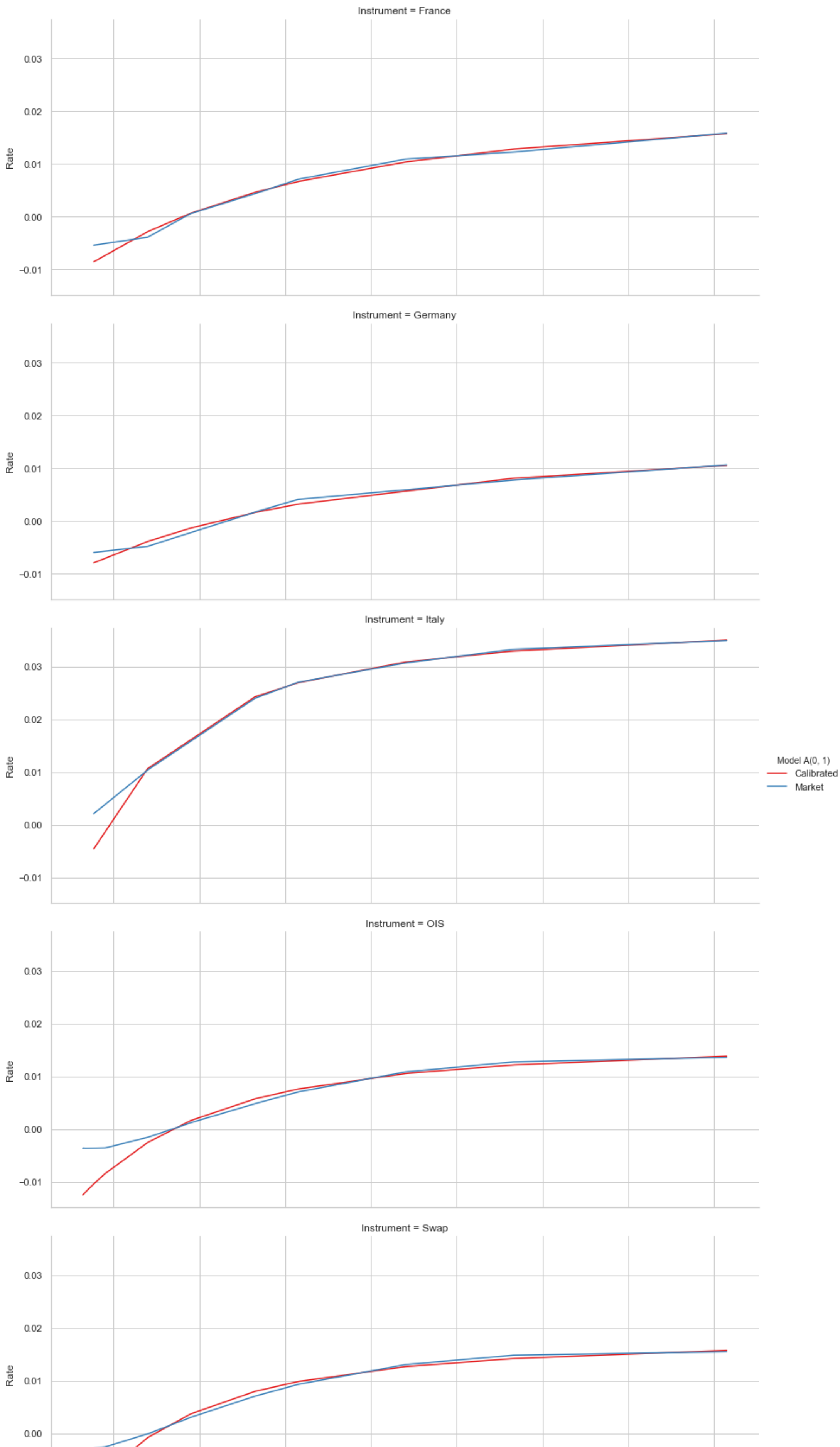
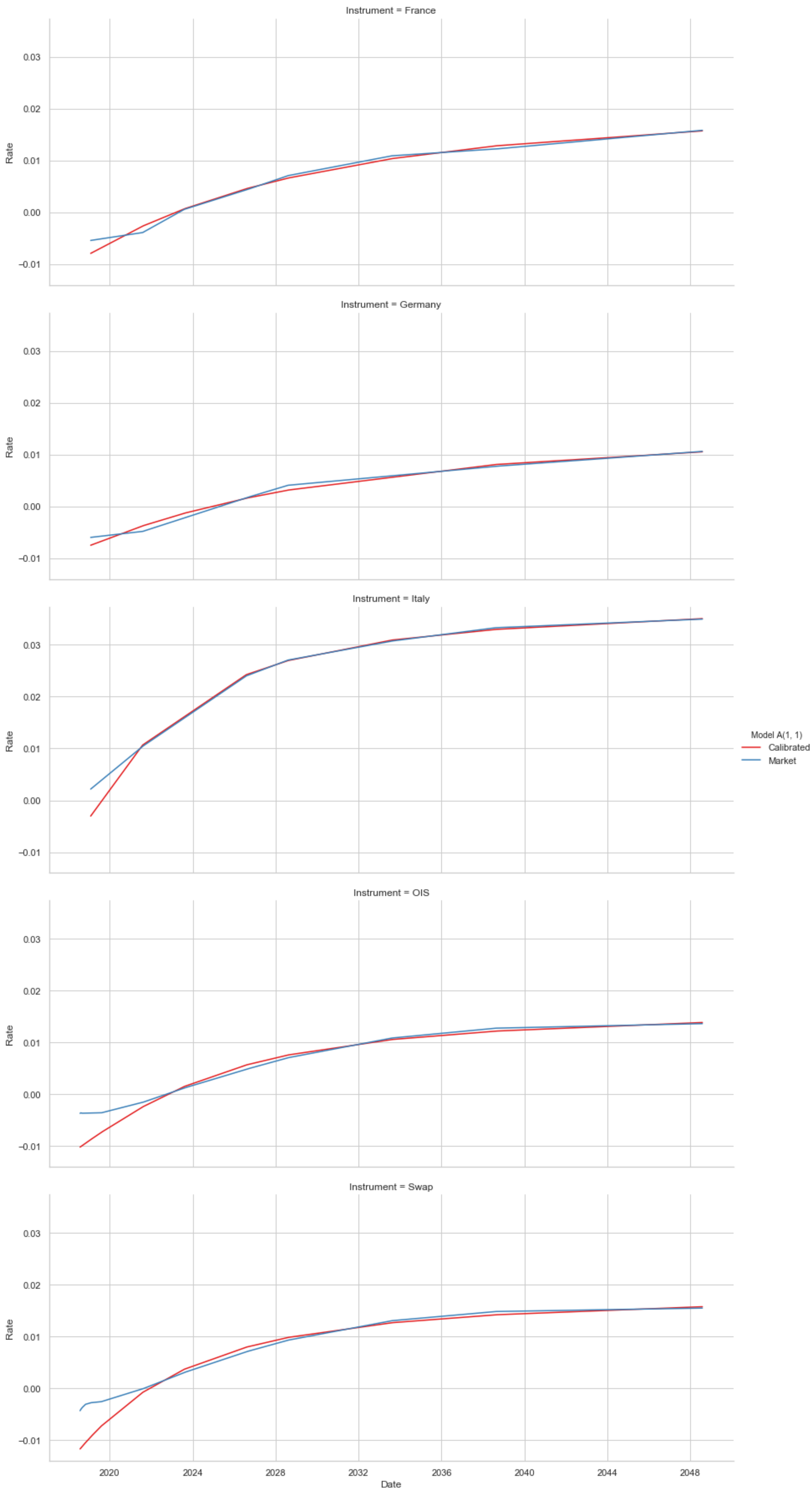


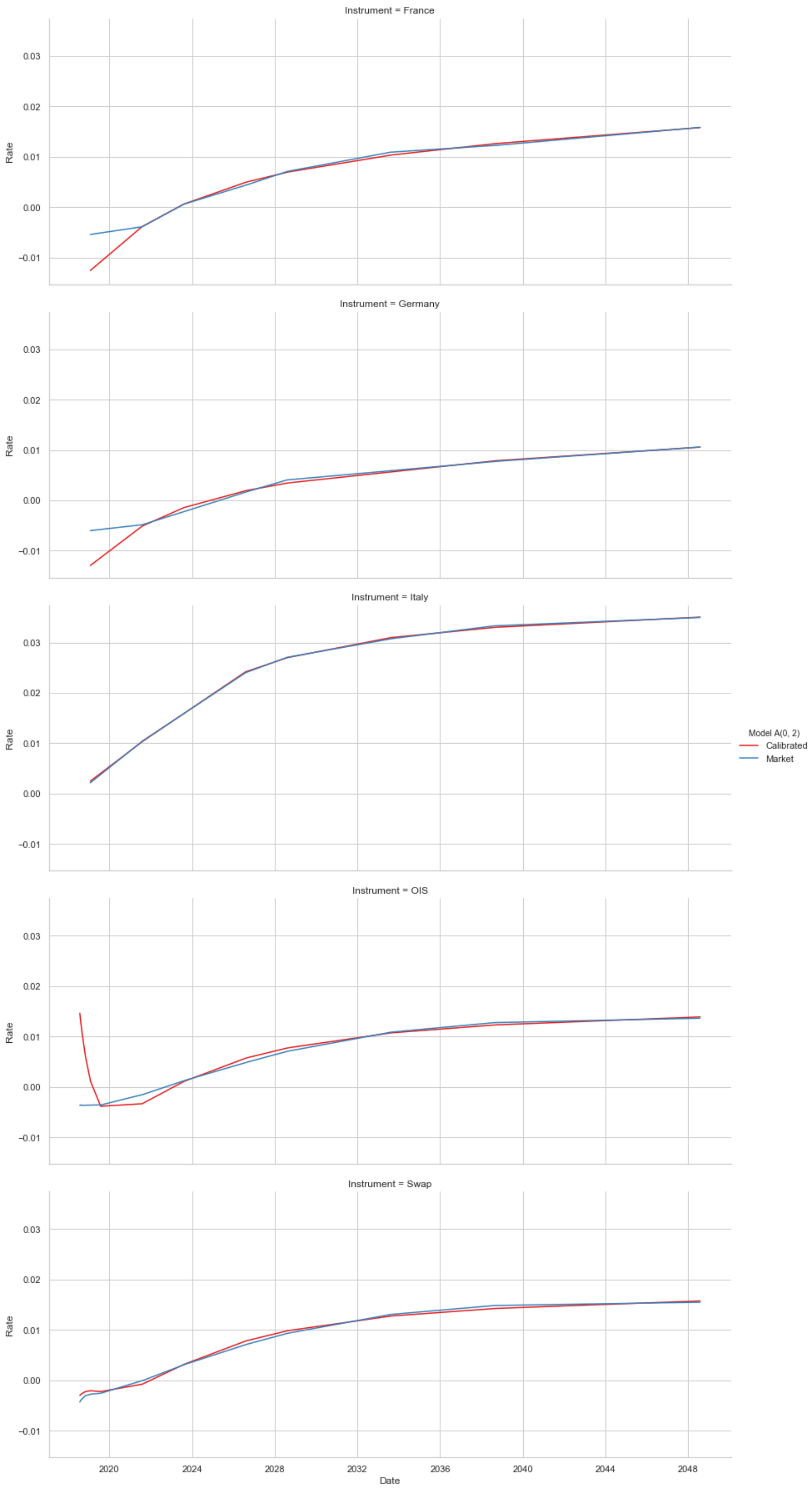
Figure B.4: Theoretical zero coupon bond prices inferred from the rate data

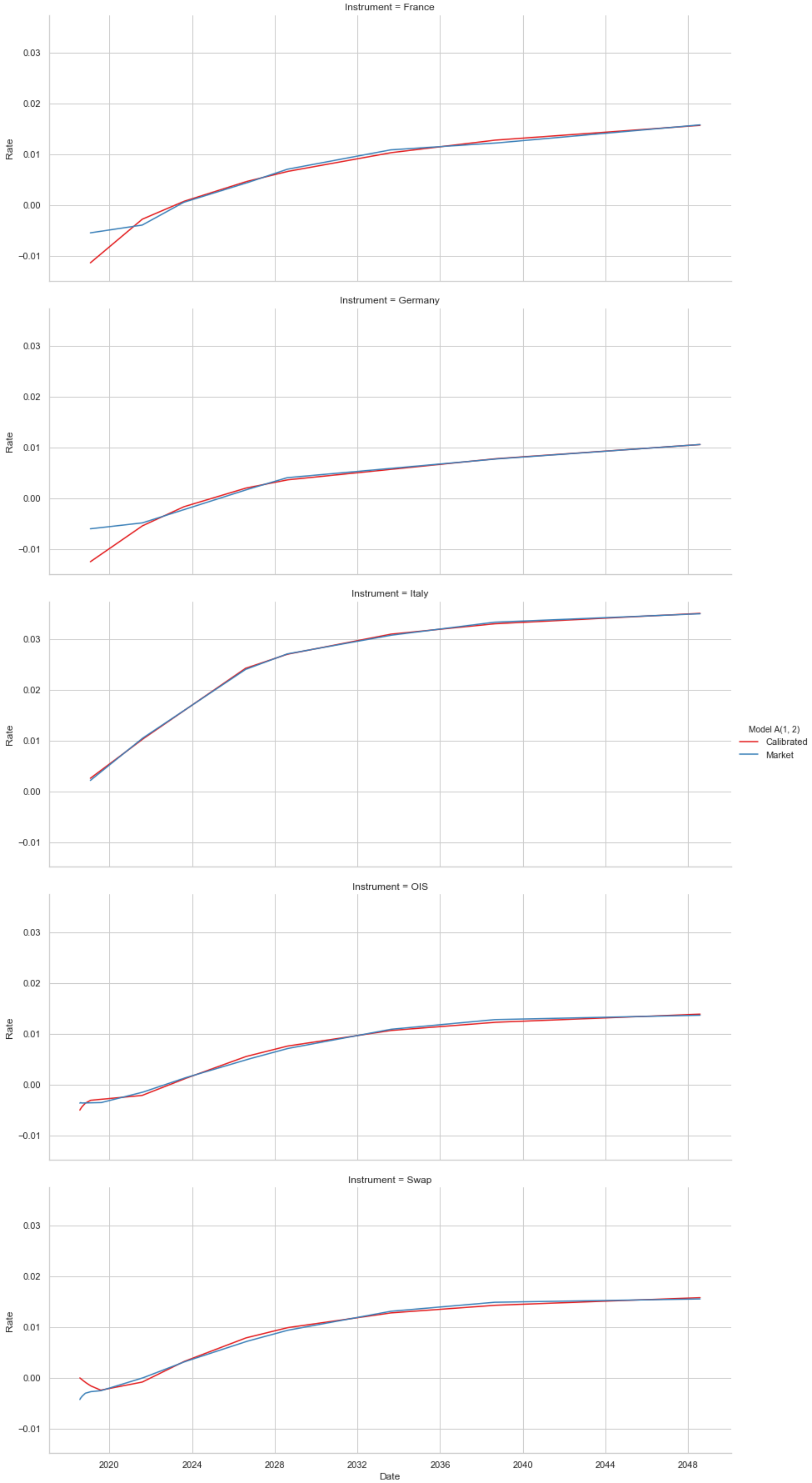
B.2. Comparison of actual inferred rates and calibrated model prices without default risk

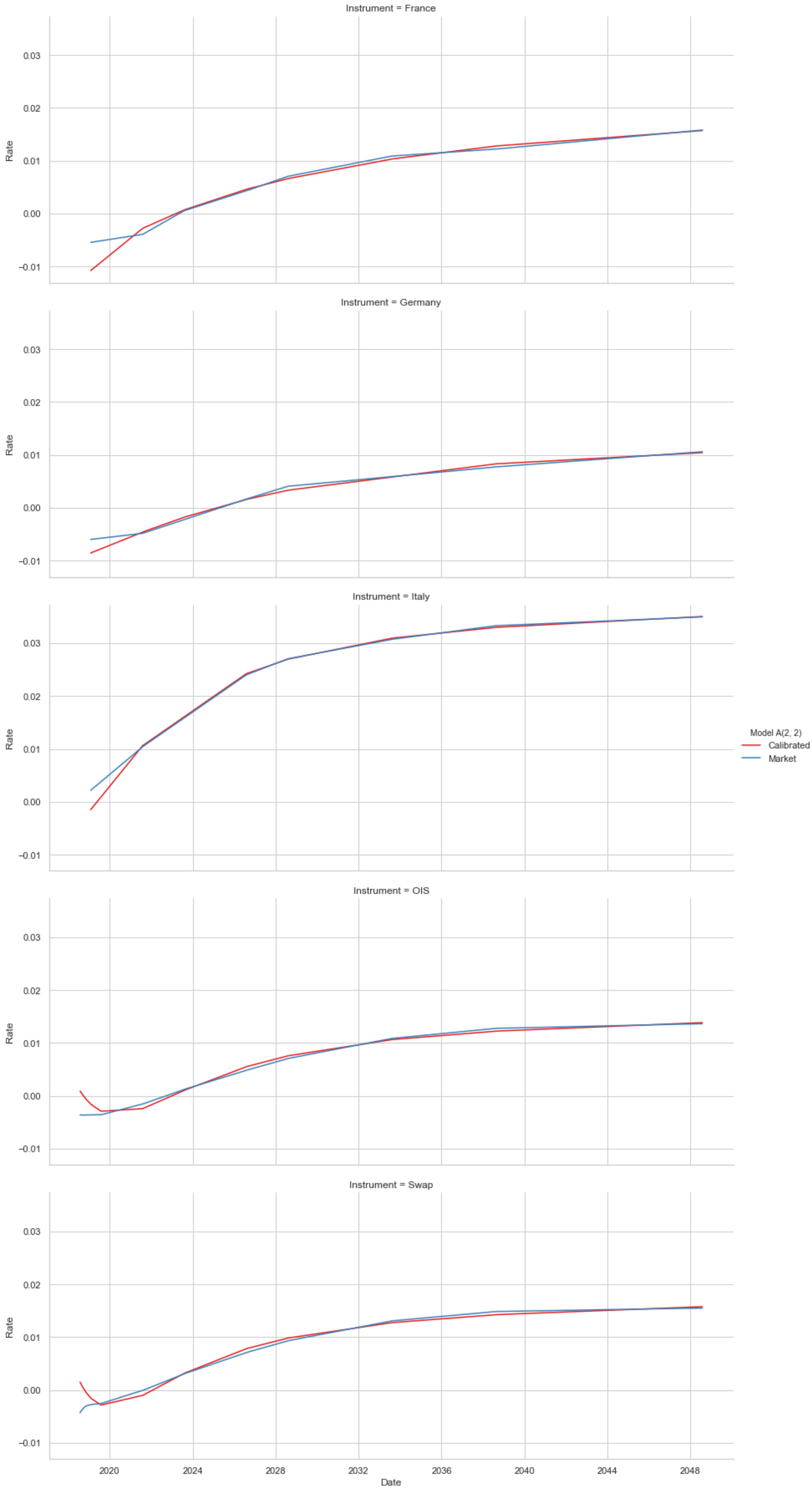




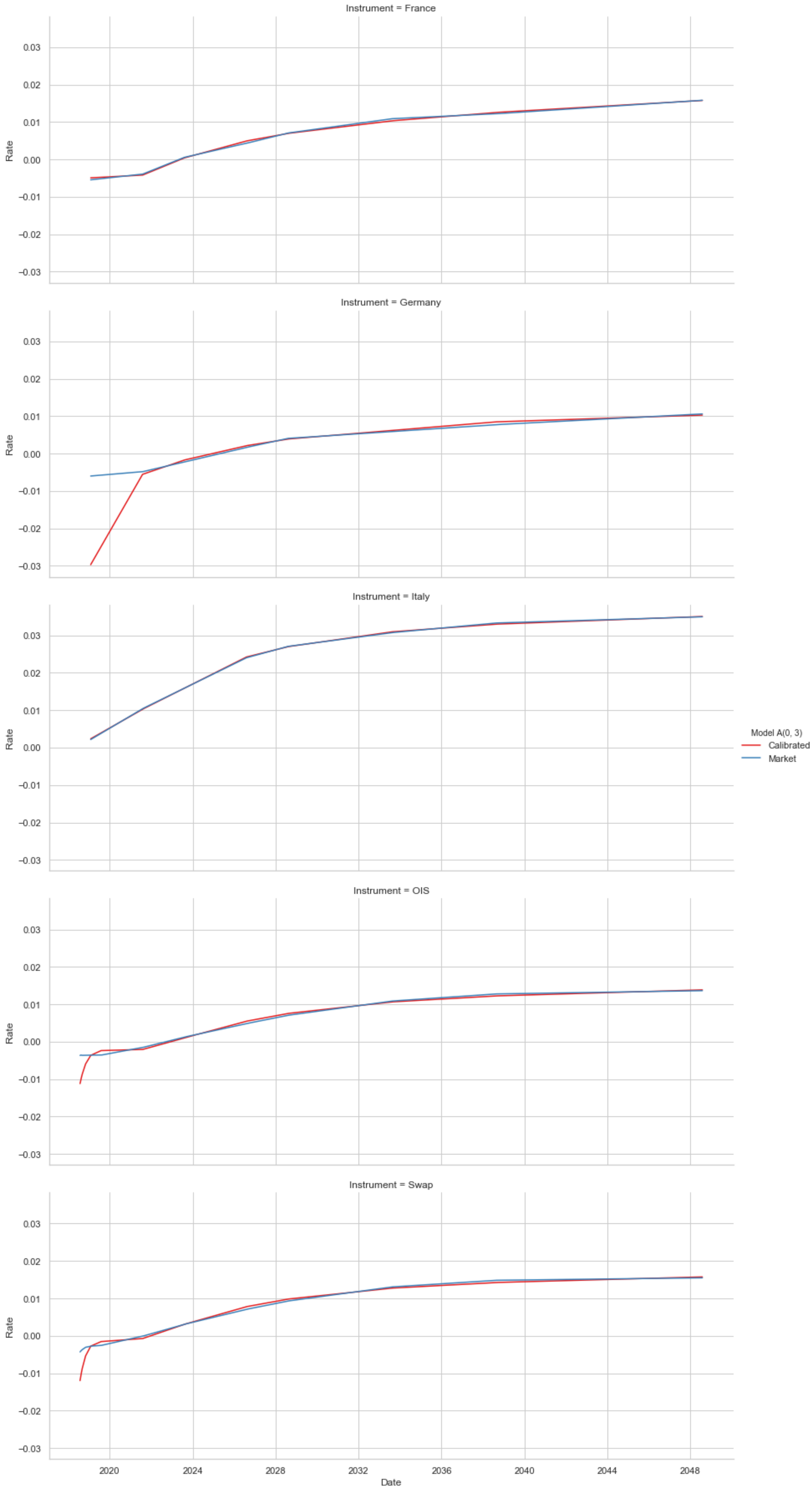
Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
Date
2020
2024
2028
2032
2036
2040
2044
2048
Model A(1, 1)
Calibrated
Market


Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
Model A(0, 2)
Calibrated
Market
2020
2024
2028
2032
2036
2040
2044
2048
Date

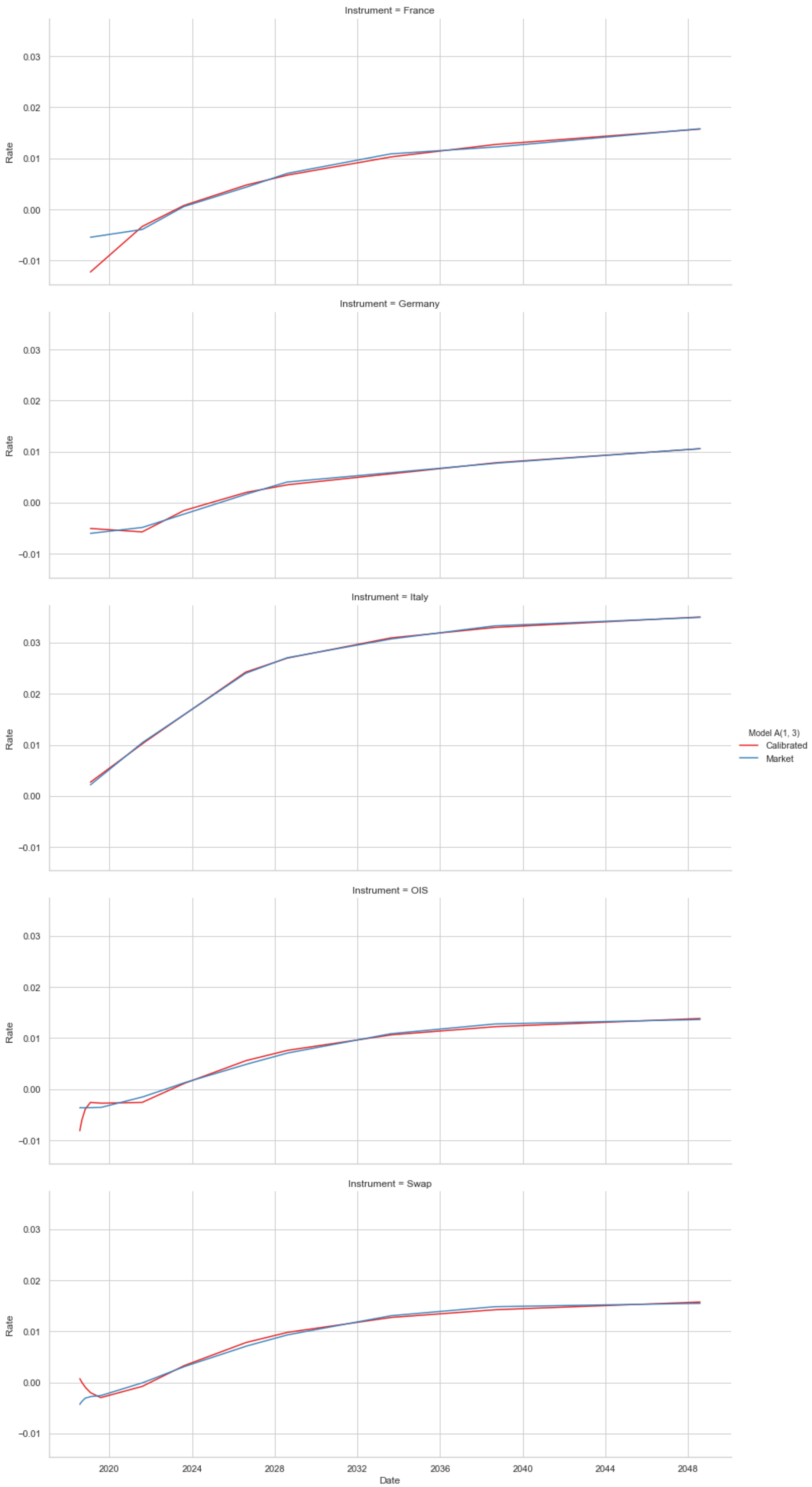

Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
2020
2024
2028
2032
2036
2040
2044
2048
Date
Model A(1, 2)
Calibrated
Market

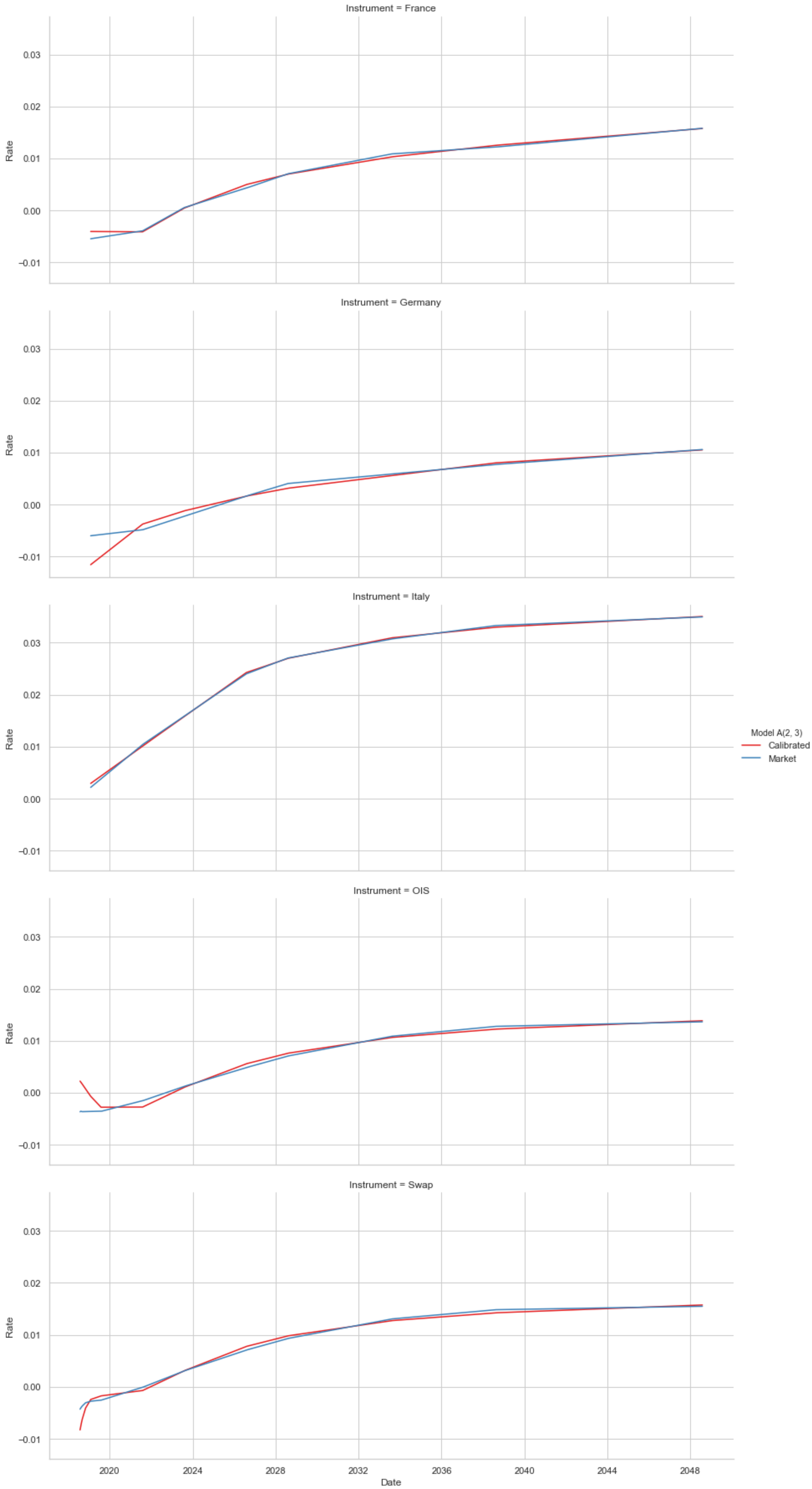


Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
2020
2024
2028
2032
2036
2040
2044
2048
Date
Model A(2, 2)
Calibrated
Market

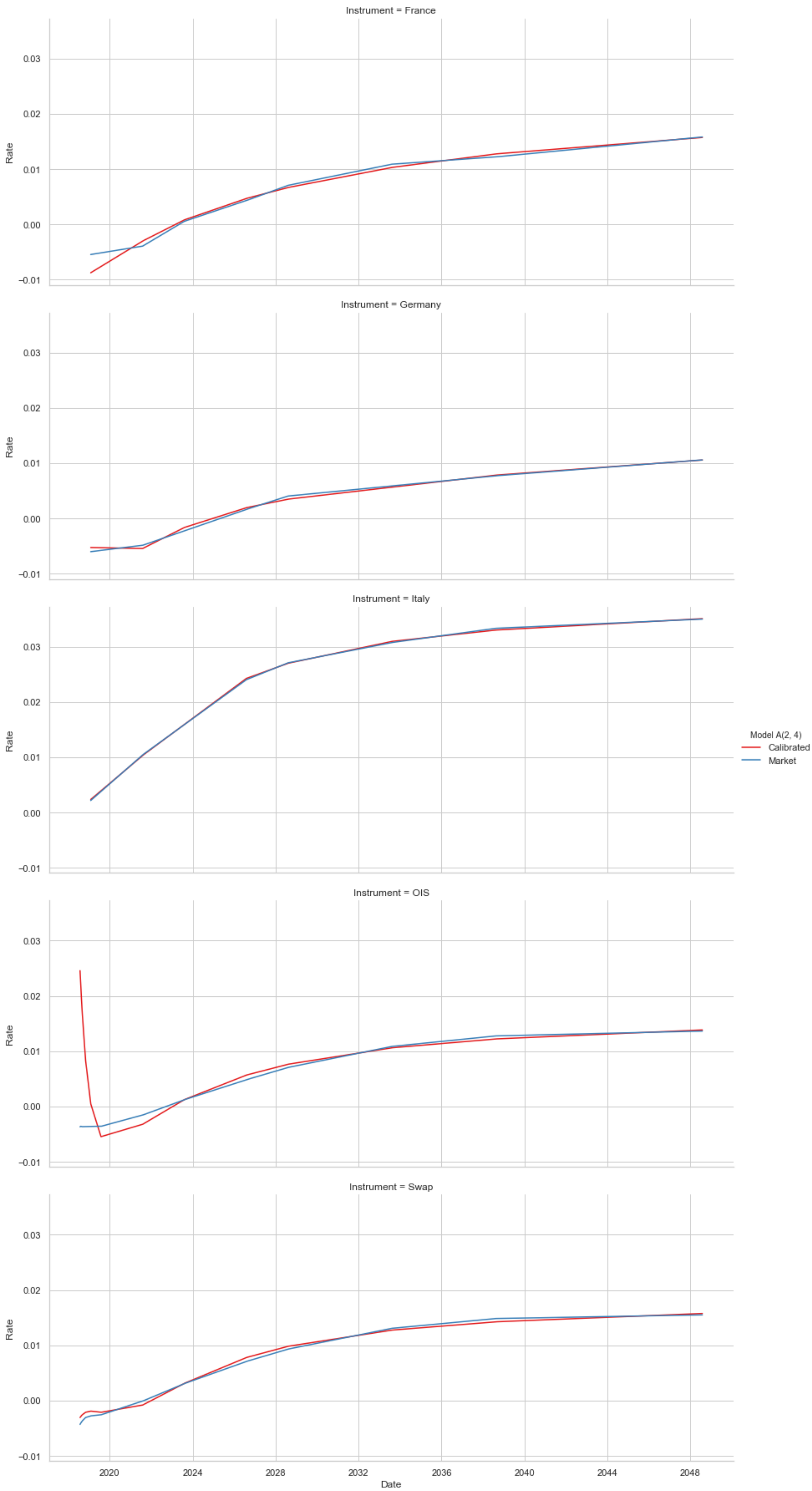


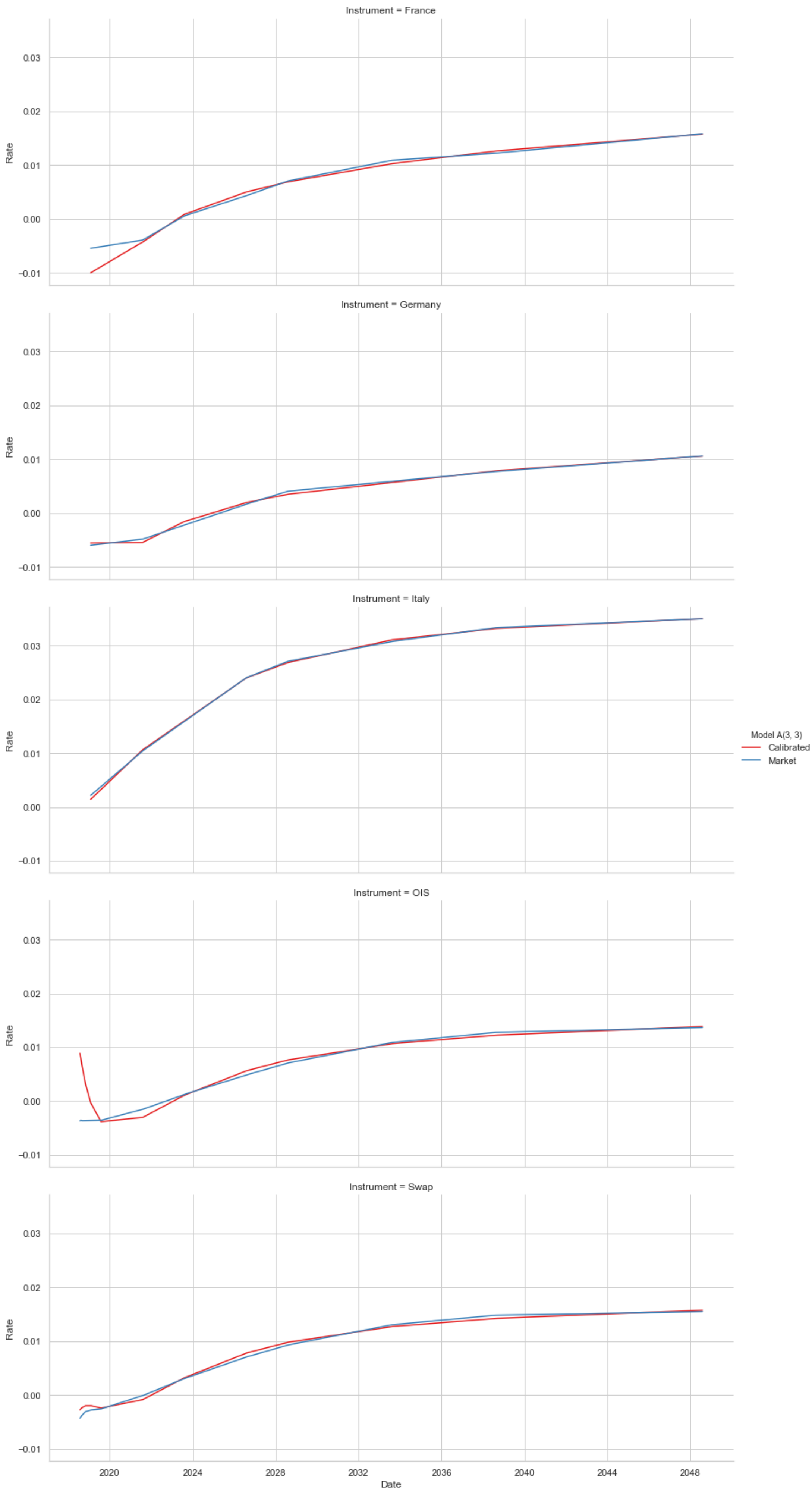
Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
2020
2024
2028
2032
2036
2040
2044
2048
Date
Model A(0, 3)
Calibrated
Market


Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
2020
2024
2028
2032
2036
2040
2044
2048
Date
Model A(1, 3)
Calibrated
Market

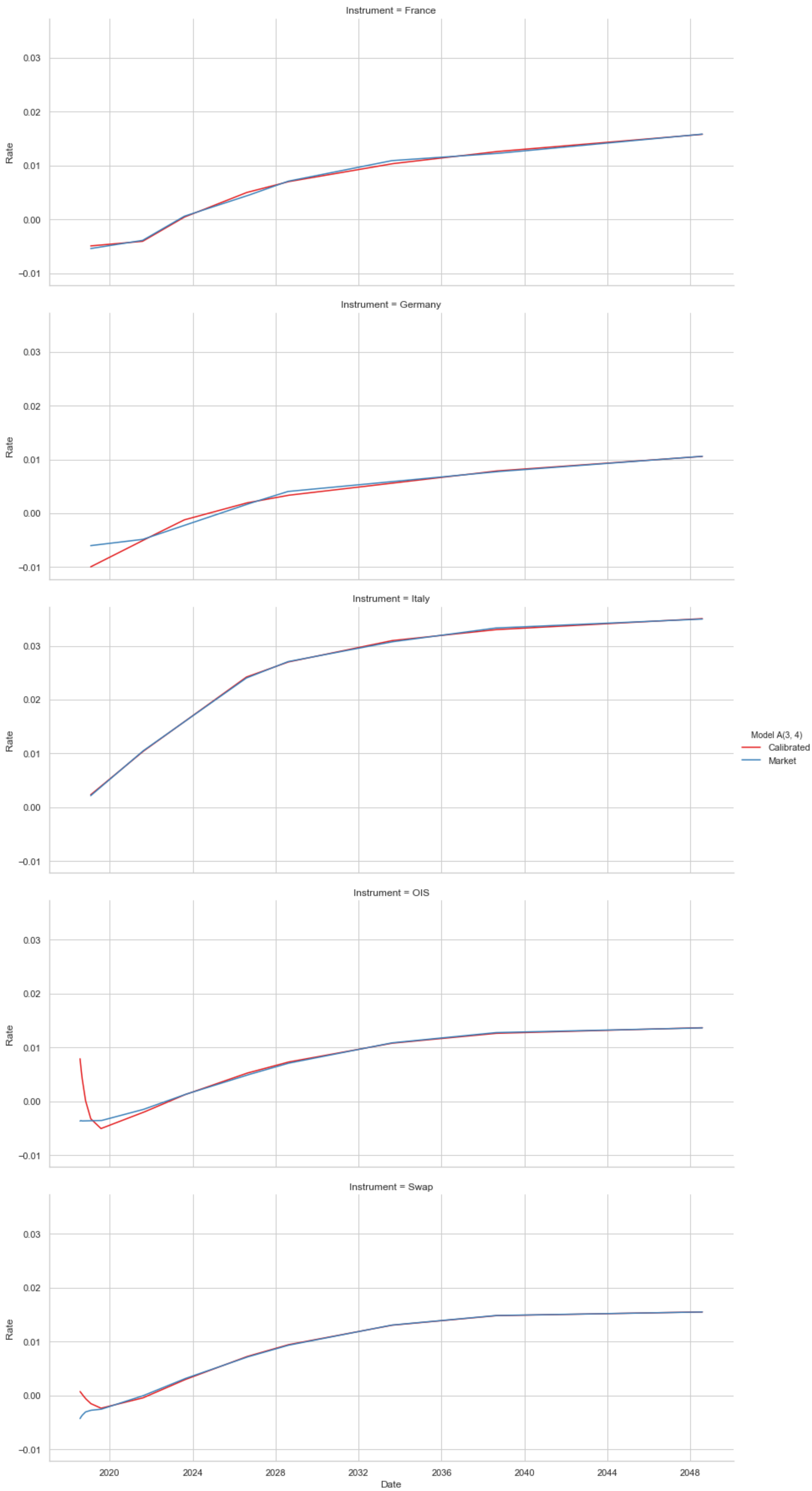


Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
Model A(2, 3)
Calibrated
Market
2020
2024
2028
2032
2036
2040
2044
2048
Date

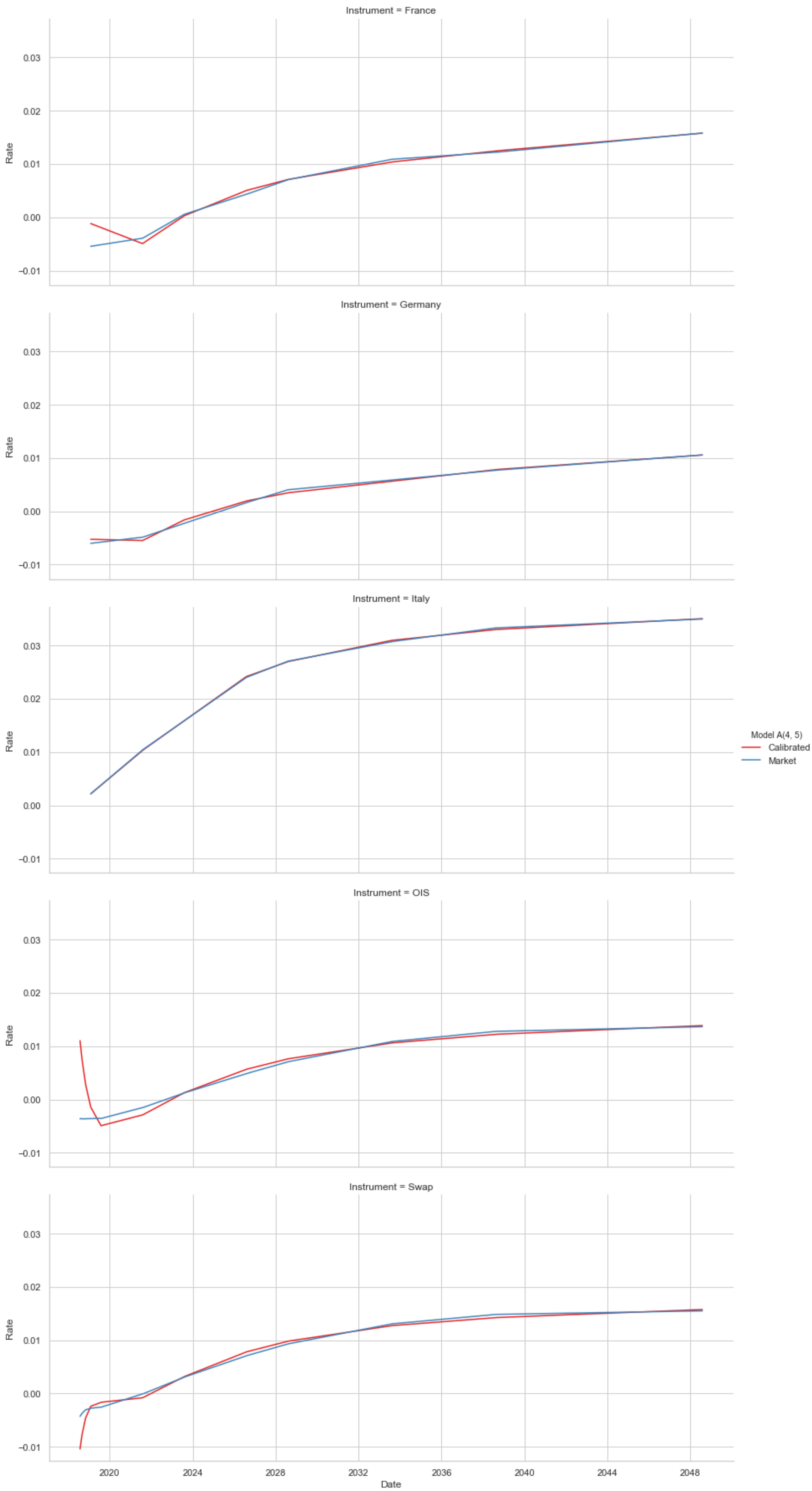

Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
Date
Model A(2, 4)
Calibrated
Market



Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
2020
2024
2028
2032
2036
2040
2044
2048
Date
Model A(3, 3)
Calibrated
Market



Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
2020
2024
2028
2032
2036
2040
2044
2048
Date
Model A(3, 4)
Calibrated
Market


Instrument = France
Instrument = Germany
Instrument = Italy
Instrument = OIS
Instrument = Swap
Rate
0.03
0.02
0.01
0.00
−0.01
Date
2020
2024
2028
2032
2036
2040
2044
2048
Model A(4, 5)
Calibrated
Market

B.3. Comparison of different calibrations between models without default risk

Figure B.17: Comparison of alternative calibrations for single factor models

Figure B.18: Comparison of alternative calibrations for 2-factor models

Figure B.19: Comparison of alternative calibrations for multifactor models

B.4. Model parameters for affine models without default risk

Table B.1: Overnight index swap

	0, 1	1, 1	0, 2	1, 2	2, 2	1, 3	0, 3	2, 3	3, 3	3, 4	2, 4	4, 5
X_0		0.001		0.001	[0.001, 0.08]	0.002		[0.001, 0.079]	[0.017, 0.007, 0.031]	[0.02, 0.084, 0.072]	[0.019, 0.03]	[0.041, 0.058, 0.038, 0.002]
Y_0	-0.030		[-0.046, 0.04]	0.021		[-0.08, 0.1]	[-0.1, 0.062, -0.022]	-0.054		-0.1	[-0.02, 0.09]	-0.038
α		0.057		0.338	[0.444, 0.703]	0.396		[2.654, 1.373]	[2.173, 0.185, 1.918]	[1.827, 2.999, 1.103]	[0.308, 1.782]	[3.0, 2.988, 2.992, 3.0]
δ	0.017	-0.011	0.02	-0.027	-0.08	-0.031	0.049	-0.023	-0.046	-0.068	-0.095	-0.089
k	0.287		[0.387, 2.362]	0.894		[2.539, 1.731]	[0.483, 0.949, 2.998]	0.404		0.059	[1.317, 2.993]	0.342
ν	0.001		[0.029, 0.212]	0.250		[0.068, 0.246]	[0.001, 0.237, 0.002]	0.013		0.024	[0.176, 0.001]	0.001
ρ			-0.566			-0.99	[0.872, -0.011, -0.19]				-0.99	
σ		0.240		0.233	[0.118, 0.247]	0.001		[0.001, 0.133]	[0.185, 0.249, 0.001]	[0.046, 0.25, 0.245]	[0.097, 0.055]	[0.159, 0.249, 0.249, 0.25]
θ		0.100		0.100	[0.1, 0.0]	0.055		[0.018, 0.023]	[0.0, 0.075, 0.016]	[0.085, 0.023, 0.05]	[0.057, 0.066]	[0.084, 0.0, 0.0, 0.021]

Table B.2: Swap

	0, 1	1, 1	0, 2	1, 2	2, 2	1, 3	0, 3	2, 3	3, 3	3, 4	2, 4	4, 5
X_0		0.001		0.001	[0.012, 0.069]	0.1		[0.001, 0.094]	[0.003, 0.033, 0.061]	[0.048, 0.025, 0.002]	[0.098, 0.1]	[0.016, 0.098, 0.001, 0.05]
Y_0	-0.031		[-0.1, 0.045]	0.050		[0.002, -0.1]	[0.081, -0.1, -0.041]	-0.038		-0.006	[-0.025, -0.1]	-0.076
α		0.279		0.381	[0.328, 0.92]	0.734		[0.388, 0.981]	[0.291, 2.303, 1.296]	[0.001, 2.249, 0.054]	[1.356, 0.905]	[0.306, 1.746, 2.521, 1.534]
δ	0.019	-0.013	0.052	-0.051	-0.079	-0.002	0.048	-0.065	-0.1	-0.068	-0.076	-0.1
k	0.326		[0.526, 0.969]	0.888		[2.722, 0.496]	[1.04, 0.541, 2.541]	2.636		3	[2.12, 0.527]	2.336
ν	0.001		[0.001, 0.25]	0.154		[0.065, 0.001]	[0.238, 0.049, 0.211]	0.124		0.157	[0.001, 0.001]	0.041
ρ			0.456			0.811	[-0.734, 0.416, 0.312]				0.99	
σ		0.129		0.248	[0.25, 0.147]	0.001		[0.248, 0.214]	[0.25, 0.25, 0.243]	[0.025, 0.25, 0.166]	[0.152, 0.222]	[0.25, 0.101, 0.21, 0.003]
θ		0.035		0.100	[0.094, 0.023]	0.021		[0.1, 0.0]	[0.072, 0.063, 0.0]	[0.045, 0.009, 0.1]	[0.095, 0.001]	[0.093, 0.0, 0.045, 0.0]

Table B.3: Germany

	0, 1	1, 1	0, 2	1, 2	2, 2	1, 3	0, 3	2, 3	3, 3	3, 4	2, 4	4, 5
X_0		0.053		0.001	[0.016, 0.072]	0.001		[0.025, 0.068]	[0.045, 0.061, 0.001]	[0.057, 0.012, 0.1]	[0.001, 0.051]	[0.071, 0.001, 0.001, 0.075]
Y_0	-0.037		[-0.021, -0.094]	-0.055		[0.057, -0.078]	[-0.054, 0.022, -0.091]	-0.014		-0.1	[-0.054, 0.094]	-0.039
α		0.077		0.022	[0.061, 0.873]	0.004		[0.052, 1.214]	[0.008, 3.0, 0.565]	[2.183, 0.057, 3.0]	[0.004, 0.609]	[3.0, 0.005, 0.534, 2.808]
δ	0.028	-0.062	0.1	0.039	-0.098	0.031	0.055	-0.097	-0.1	-0.071	-0.091	-0.1
k	0.105		[0.428, 0.005]	0.099		[2.674, 0.58]	[0.186, 1.371, 3.0]	2.499		1.182	[1.48, 1.537]	0.855
ν	0.016		[0.001, 0.001]	0.038		[0.01, 0.133]	[0.036, 0.25, 0.189]	0.023		0.246	[0.03, 0.086]	0.23
ρ			0.641			0.558	[0.422, -0.924, -0.044]				-0.184	
σ		0.062		0.001	[0.133, 0.15]	0.001		[0.073, 0.075]	[0.001, 0.181, 0.072]	[0.001, 0.001, 0.12]	[0.001, 0.13]	[0.002, 0.001, 0.225, 0.204]
θ		0.098		0.100	[0.09, 0.071]	0.1		[0.077, 0.066]	[0.1, 0.027, 0.035]	[0.001, 0.031, 0.081]	[0.095, 0.1]	[0.05, 0.085, 0.057, 0.039]

Table B.4: France

	0, 1	1, 1	0, 2	1, 2	2, 2	1, 3	0, 3	2, 3	3, 3	3, 4	2, 4	4, 5
X_0		0.016		0.019	[0.02, 0.064]	0.067		[0.06, 0.002]	[0.095, 0.001, 0.009]	[0.002, 0.012, 0.043]	[0.093, 0.07]	[0.085, 0.062, 0.001, 0.049]
Y_0	-0.049		[-0.088, -0.027]	-0.010		[-0.1, 0.084]	[-0.036, 0.028, -0.087]	0.034		0.032	[0.039, -0.1]	-0.099
α		0.159		0.148	[2.999, 0.131]	0.1		[0.469, 0.005]	[2.411, 1.31, 0.064]	[0.012, 0.545, 2.65]	[2.999, 0.134]	[1.868, 0.018, 1.162, 2.772]
δ	0.039	-0.025	0.1	-0.026	-0.1	-0.066	0.1	-0.09	-0.1	-0.087	-0.1	-0.076
k	0.115		[0.005, 0.348]	2.492		[1.241, 1.384]	[0.445, 2.866, 0.005]	2.298		0.963	[2.995, 2.199]	0.047
ν	0.021		[0.001, 0.001]	0.097		[0.026, 0.137]	[0.011, 0.21, 0.001]	0.001		0.177	[0.177, 0.189]	0.002
ρ			-0.99			-0.374	[-0.836, -0.894, 0.503]				-0.572	
σ		0.006		0.011	[0.179, 0.046]	0.038		[0.007, 0.001]	[0.1, 0.116, 0.039]	[0.023, 0.218, 0.195]	[0.042, 0.01]	[0.085, 0.085, 0.234, 0.005]
θ		0.047		0.049	[0.028, 0.1]	0.1		[0.1, 0.1]	[0.0, 0.095, 0.035]	[0.049, 0.099, 0.021]	[0.027, 0.099]	[0.004, 0.049, 0.1, 0.008]

Table B.5: Italy

	0, 1	1, 1	0, 2	1, 2	2, 2	1, 3	0, 3	2, 3	3, 3	3, 4	2, 4	4, 5
X_0		0.001		0.016	[0.001, 0.051]	0.06		[0.004, 0.093]	[0.014, 0.018, 0.02]	[0.1, 0.001, 0.033]	[0.078, 0.004]	[0.024, 0.003, 0.04, 0.08]
Y_0	-0.048		[-0.1, 0.064]	0.024		[0.022, -0.061]	[0.1, -0.074, -0.065]	-0.055		-0.077	[-0.026, 0.025]	-0.075
α		0.213		0.314	[0.433, 0.647]	1.893		[1.978, 3.0]	[0.328, 1.651, 0.001]	[1.503, 2.256, 1.875]	[1.343, 0.413]	[2.978, 0.376, 0.998, 0.822]
δ	0.039	-0.008	0.039	-0.032	-0.056	-0.012	0.039	-0.025	-0.05	-0.058	-0.076	-0.071
k	0.378		[0.511, 0.904]	2.483		[2.536, 0.43]	[2.141, 2.818, 0.448]	0.418		0.478	[0.906, 1.684]	0.789
ν	0.001		[0.001, 0.001]	0.114		[0.063, 0.016]	[0.012, 0.007, 0.001]	0.001		0.009	[0.046, 0.057]	0.102
ρ			-0.886			0.99	[0.868, 0.508, 0.868]				0.243	
σ		0.250		0.216	[0.137, 0.142]	0.16		[0.029, 0.001]	[0.09, 0.247, 0.018]	[0.249, 0.112, 0.197]	[0.12, 0.001]	[0.233, 0.247, 0.016, 0.25]
θ		0.069		0.086	[0.099, 0.0]	0.053		[0.029, 0.035]	[0.07, 0.004, 0.006]	[0.0, 0.046, 0.052]	[0.061, 0.057]	[0.028, 0.1, 0.001, 0.006]

B.5. Comparison of parameters for affine models without default risk between different calibration attempts

OIS		Original	Alternative
$A(0,1)+$	Y_0	-0.03	-0.03
	δ	0.017	0.017
	k	0.287	0.286
	ν	0.001	0.001
$A(0,2)+$	Y_0	[-0.046, 0.04]	[0.04, -0.1]
	δ	0.02	0.055
	k	[0.387, 2.362]	[0.864, 0.481]
	ν	[0.029, 0.212]	[0.25, 0.013]
	ρ	-0.566	-0.562
$A(1,1)+$	X_0	0.001	0.001
	α	0.057	0.057
	δ	-0.011	-0.011
	σ	0.24	0.24
	θ	0.1	0.1
$A(2,2)+$	X_0	[0.001, 0.08]	[0.059, 0.022]
	α	[0.444, 0.703]	[0.807, 0.416]
	δ	-0.08	-0.079
	σ	[0.118, 0.247]	[0.12, 0.103]
	θ	[0.1, 0.0]	[0.0, 0.099]
$A(3,3)+$	X_0	[0.017, 0.007, 0.031]	[0.037, 0.03, 0.001]
	α	[2.173, 0.185, 1.918]	[2.989, 2.964, 0.111]
	δ	-0.046	-0.06
	σ	[0.185, 0.249, 0.001]	[0.25, 0.175, 0.25]
	θ	[0.0, 0.075, 0.016]	[0.023, 0.021, 0.073]
$A(3,4)+$	X_0	[0.02, 0.084, 0.072]	[0.004, 0.07, 0.079]
	Y_0	-0.1	-0.099
	α	[1.827, 2.999, 1.103]	[0.252, 2.995, 3.0]
	δ	-0.068	-0.044
	k	0.059	2.999
	ν	0.024	0.019
	σ	[0.046, 0.25, 0.245]	[0.181, 0.001, 0.25]
	θ	[0.085, 0.023, 0.05]	[0.05, 0.016, 0.004]

Swap		Original	Alternative
$A(0,1)+$	Y_0	-0.031	-0.031
	δ	0.019	0.019
	k	0.326	0.326
	ν	0.001	0.001
$A(0,2)+$	Y_0	[-0.1, 0.045]	[0.055, -0.07]
	δ	0.052	0.019
	k	[0.526, 0.969]	[1.062, 0.488]
	ν	[0.001, 0.25]	[0.018, 0.001]
	ρ	0.456	-0.643
$A(1,1)+$	X_0	0.001	0.001
	α	0.279	0.096
	δ	-0.013	-0.01
	σ	0.129	0.25
	θ	0.035	0.071
$A(2,2)+$	X_0	[0.012, 0.069]	[0.079, 0.02]
	α	[0.328, 0.92]	[2.115, 0.418]
	δ	-0.079	-0.089
	σ	[0.25, 0.147]	[0.137, 0.021]
	θ	[0.094, 0.023]	[0.043, 0.065]
$A(3,3)+$	X_0	[0.003, 0.033, 0.061]	[0.054, 0.027, 0.001]
	α	[0.291, 2.303, 1.296]	[0.736, 0.007, 0.079]
	δ	-0.1	-0.085
	σ	[0.25, 0.25, 0.243]	[0.198, 0.048, 0.121]
	θ	[0.072, 0.063, 0.0]	[0.047, 0.0, 0.075]
$A(3,4)+$	X_0	[0.048, 0.025, 0.002]	[0.001, 0.1, 0.001]
	Y_0	-0.006	-0.009
	α	[0.001, 2.249, 0.054]	[2.012, 1.052, 0.381]
	δ	-0.068	-0.1
	k	3	1.625
	ν	0.157	0.155
	σ	[0.025, 0.25, 0.166]	[0.105, 0.164, 0.25]
	θ	[0.045, 0.009, 0.1]	[0.042, 0.0, 0.096]

Germany		Original	Alternative
$A(0,1)+$	Y_0	-0.037	-0.037
	δ	0.028	0.028
	k	0.105	0.105
	ν	0.016	0.016
$A(0,2)+$	Y_0	[-0.021, -0.094]	[-0.022, -0.059]
	δ	0.1	0.025
	k	[0.428, 0.005]	[0.063, 2.854]
	ν	[0.001, 0.001]	[0.001, 0.219]
	ρ	0.641	0.99
$A(1,1)+$	X_0	0.053	0.027
	α	0.077	0.139
	δ	-0.062	-0.036
	σ	0.062	0.009
	θ	0.098	0.052
$A(2,2)+$	X_0	[0.016, 0.072]	[0.012, 0.002]
	α	[0.061, 0.873]	[0.005, 0.304]
	δ	-0.098	-0.028
	σ	[0.133, 0.15]	[0.001, 0.221]
	θ	[0.09, 0.071]	[0.1, 0.027]
$A(3,3)+$	X_0	[0.045, 0.061, 0.001]	[0.021, 0.036, 0.023]
	α	[0.008, 3.0, 0.565]	[0.175, 0.001, 1.515]
	δ	-0.1	-0.086
	σ	[0.001, 0.181, 0.072]	[0.033, 0.223, 0.25]
	θ	[0.1, 0.027, 0.035]	[0.099, 0.087, 0.005]
$A(3,4)+$	X_0	[0.057, 0.012, 0.1]	[0.001, 0.082, 0.099]
	Y_0	-0.1	-0.099
	α	[2.183, 0.057, 3.0]	[0.512, 0.613, 0.624]
	δ	-0.071	-0.1
	k	1.182	0.005
	ν	0.246	0.001
	σ	[0.001, 0.001, 0.12]	[0.242, 0.242, 0.25]
	θ	[0.001, 0.031, 0.081]	[0.1, 0.058, 0.064]

France		Original	Alternative
$A(0,1)+$	Y_0	-0.049	-0.049
	δ	0.039	0.039
	k	0.115	0.115
	ν	0.021	0.021
$A(0,2)+$	Y_0	[-0.088, -0.027]	[-0.027, -0.088]
	δ	0.1	0.1
	k	[0.005, 0.348]	[0.349, 0.005]
	ν	[0.001, 0.001]	[0.003, 0.001]
	ρ	-0.99	-0.65
$A(1,1)+$	X_0	0.016	0.053
	α	0.159	0.107
	δ	-0.025	-0.062
	σ	0.006	0.071
	θ	0.047	0.1
$A(2,2)+$	X_0	[0.02, 0.064]	[0.001, 0.001]
	α	[2.999, 0.131]	[0.185, 0.004]
	δ	-0.1	-0.016
	σ	[0.179, 0.046]	[0.231, 0.001]
	θ	[0.028, 0.1]	[0.041, 0.1]
$A(3,3)+$	X_0	[0.095, 0.001, 0.009]	[0.001, 0.002, 0.076]
	α	[2.411, 1.31, 0.064]	[0.004, 0.489, 0.725]
	δ	-0.1	-0.083
	σ	[0.1, 0.116, 0.039]	[0.001, 0.143, 0.14]
	θ	[0.0, 0.095, 0.035]	[0.098, 0.098, 0.0]
$A(3,4)+$	X_0	[0.002, 0.012, 0.043]	[0.016, 0.069, 0.001]
	Y_0	0.032	-0.004
	α	[0.012, 0.545, 2.65]	[0.005, 2.999, 0.39]
	δ	-0.087	-0.075
	k	0.963	2.175
	ν	0.177	0.001
	σ	[0.023, 0.218, 0.195]	[0.001, 0.222, 0.16]
	θ	[0.049, 0.099, 0.021]	[0.1, 0.034, 0.04]

Italy		Original	Alternative
$A(0,1)+$	Y_0	-0.048	-0.048
	δ	0.039	0.039
	k	0.378	0.378
	ν	0.001	0.001
$A(0,2)+$	Y_0	[-0.1, 0.064]	[0.062, -0.1]
	δ	0.039	0.041
	k	[0.511, 0.904]	[0.919, 0.512]
	ν	[0.001, 0.001]	[0.057, 0.001]
	ρ	-0.886	0.173
$A(1,1)+$	X_0	0.001	0.001
	α	0.213	0.213
	δ	-0.008	-0.008
	σ	0.25	0.25
	θ	0.069	0.069
$A(2,2)+$	X_0	[0.001, 0.051]	[0.038, 0.011]
	α	[0.433, 0.647]	[0.005, 0.318]
	δ	-0.056	-0.055
	σ	[0.137, 0.142]	[0.021, 0.005]
	θ	[0.099, 0.0]	[0.066, 0.057]
$A(3,3)+$	X_0	[0.014, 0.018, 0.02]	[0.001, 0.046, 0.011]
	α	[0.328, 1.651, 0.001]	[0.028, 2.362, 0.385]
	δ	-0.05	-0.05
	σ	[0.09, 0.247, 0.018]	[0.25, 0.087, 0.227]
	θ	[0.07, 0.004, 0.006]	[0.076, 0.018, 0.069]
$A(3,4)+$	X_0	[0.1, 0.001, 0.033]	[0.047, 0.025, 0.086]
	Y_0	-0.077	-0.1
	α	[1.503, 2.256, 1.875]	[1.761, 2.749, 1.214]
	δ	-0.058	-0.065
	k	0.478	0.525
	ν	0.009	0.001
	σ	[0.249, 0.112, 0.197]	[0.121, 0.234, 0.016]
	θ	[0.0, 0.046, 0.052]	[0.011, 0.094, 0.0]

B.6. Actual prices and calibrated prices of default models

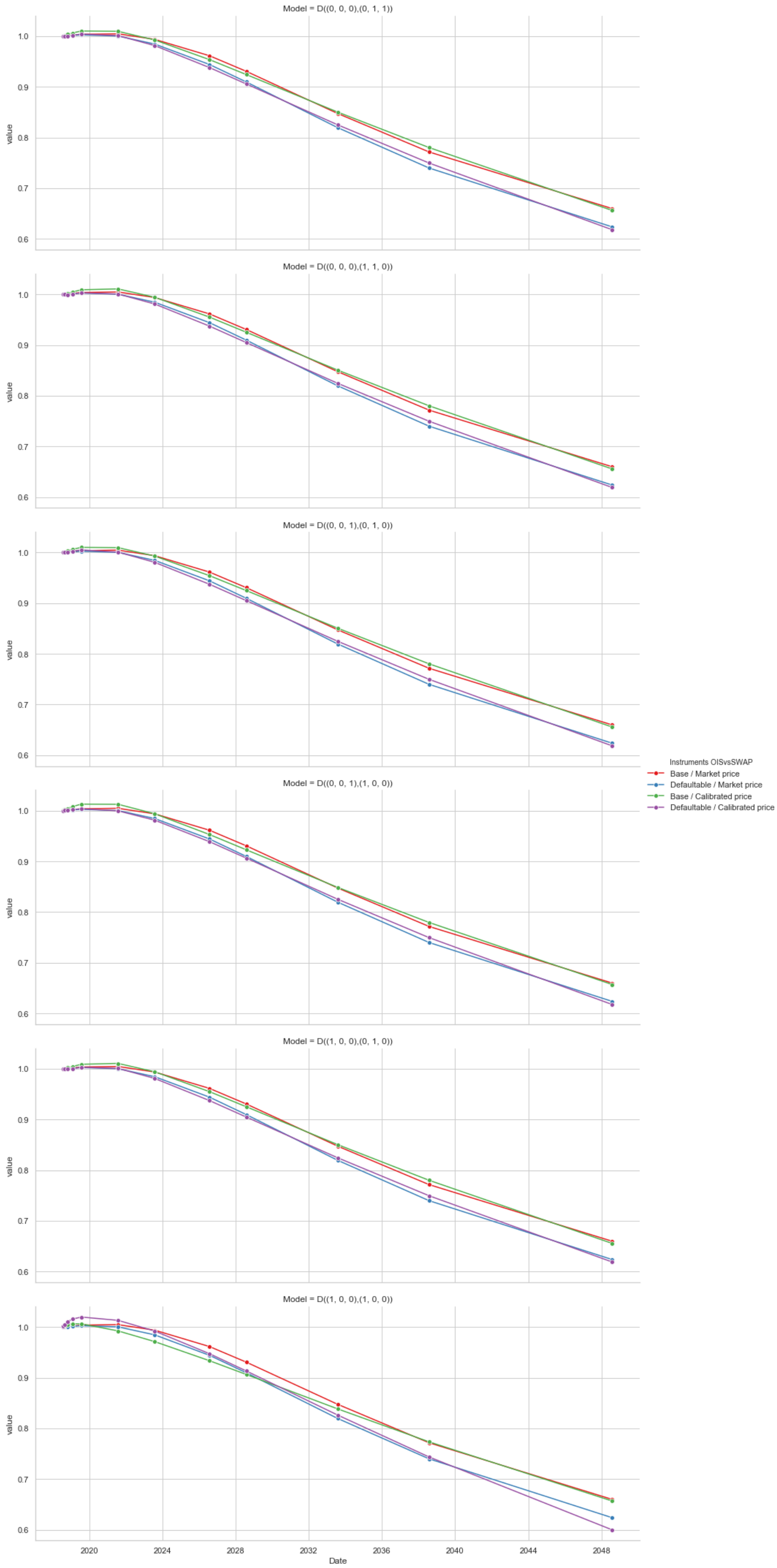


Figure B.20: Calibration and actual prices for OIS and swap rate

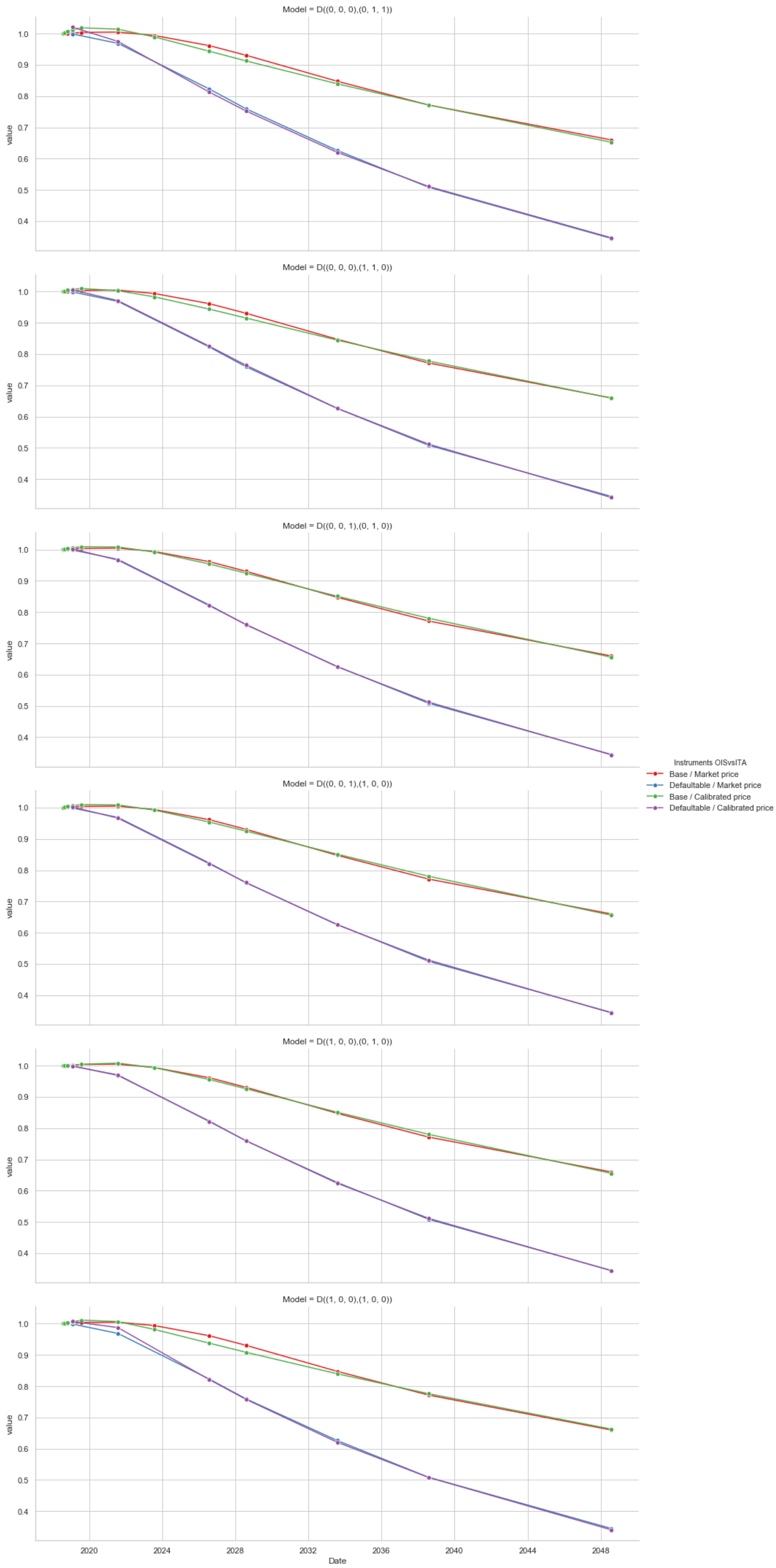


Figure B.21: Calibration and actual prices for OIS and Italy

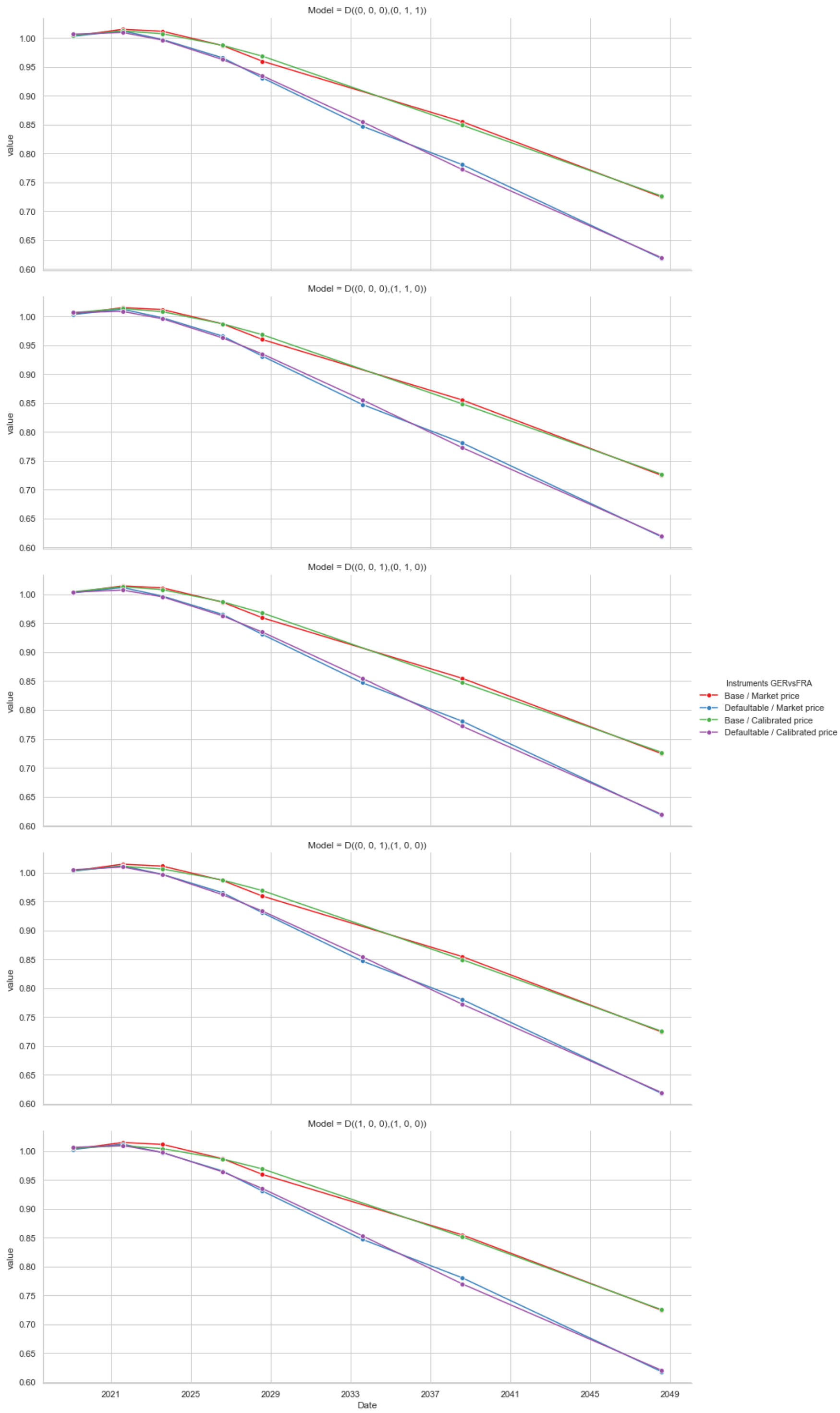


Figure B.22: Calibration and actual prices for Germany and France

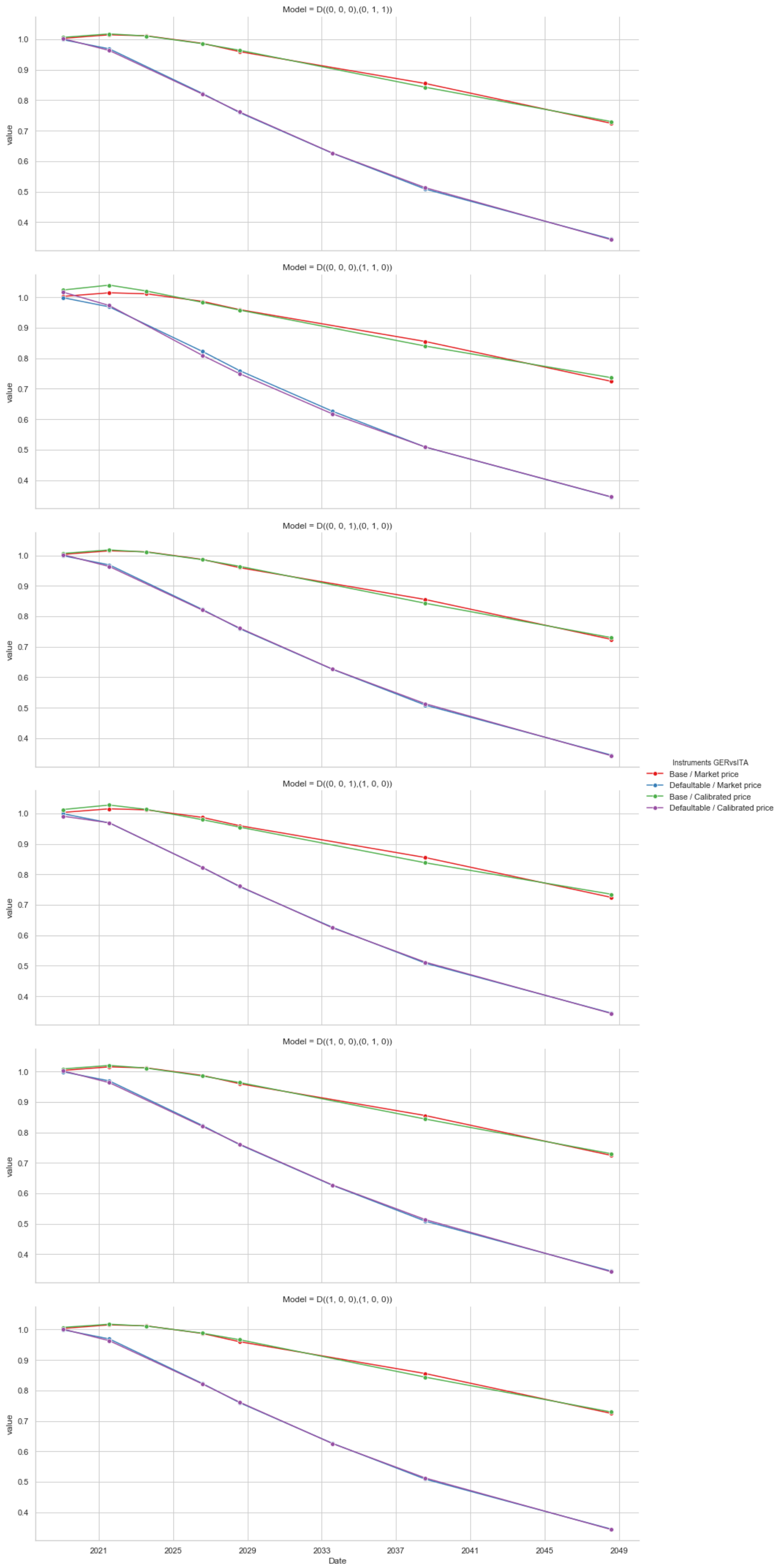


Figure B.23: Calibration and actual prices for Germany and Italy

B.7. Relative errors of default models

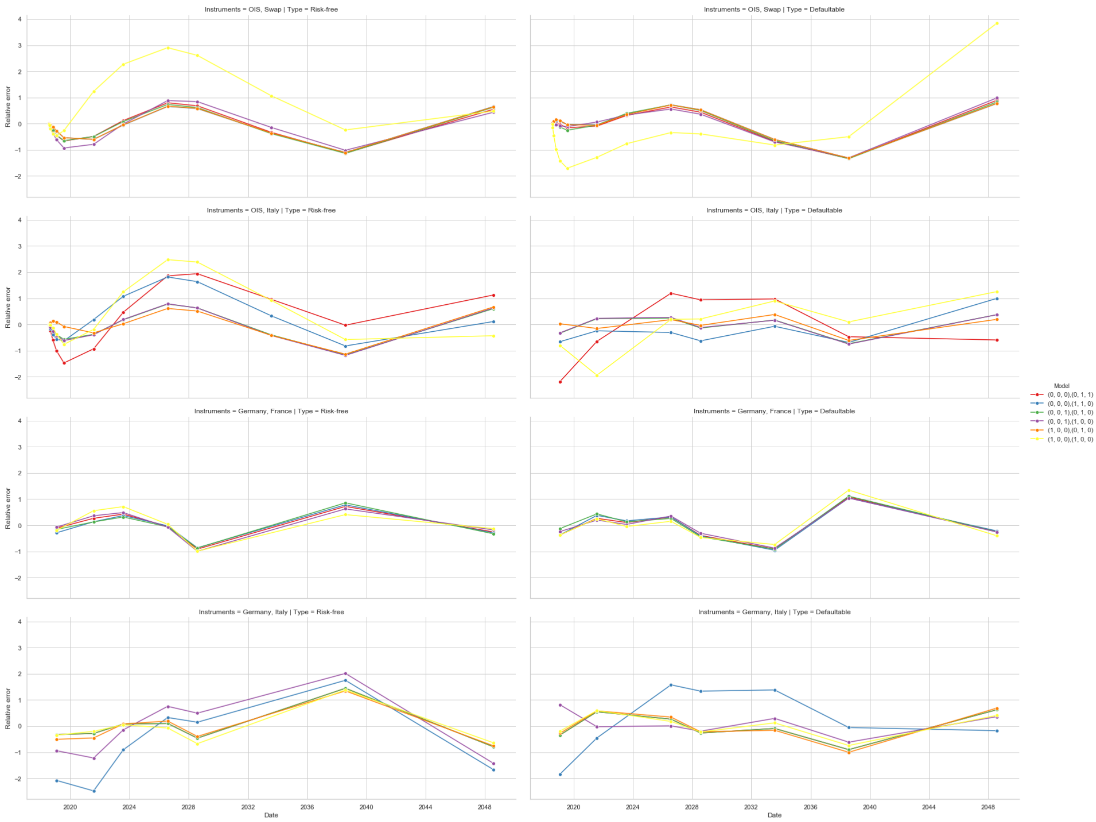


Figure B.24: Relative calibration errors for credit risk models

Figure B.25: Absolute relative calibration errors for credit risk models

B.8. Comparison of calibrated parameters for affine models with default risk

Table B.6: Overnight index swap and swap curve

	(0, 0, 1),(0, 1, 0)	(0, 0, 1),(1, 0, 0)	(1, 0, 0),(1, 0, 0)	(1, 0, 0),(0, 1, 0)	(0, 0, 0),(0, 1, 1)	(0, 0, 0),(1, 1, 0)
LGD	0.97	0.426	0.441	0.435	0.776	0.748
X_0	0.01	0.047	0.001	0.013	NaN	NaN
Y_0	-0.032	-0.036	-0.033	-0.036	[-0.032, 0.01]	[0.012, -0.036]
a_m	[0]	[0]	[1]	[1]	[]	[]
α	3	0.544	0.228	2.857	NaN	NaN
b_m	[1]	[1]	[1]	[1]	[]	[]
c_i	[1]	[1]	[1]	[1]	[1, 0]	[1, 1]
d_i	[0]	[1]	[1]	[0]	[0, 1]	[1, 0]
δ	0.017	0.017	0.005	0.017	0.017	0.017
k	0.314	0.384	2.508	0.342	[0.321, 2.727]	[2.48, 0.342]
ν	0.001	0.002	0.219	0.001	[0.007, 0.25]	[0.001, 0.001]
ρ	NaN	NaN	NaN	NaN	-0.99	0.258
σ	0.041	0.041	0.032	0.161	NaN	NaN
spread	0.002	0.002	0.002	0.002	0.004	0.002
θ	0	0	0.015	0	NaN	NaN

Table B.7: Overnight index swap and Italy

	(0, 0, 1),(0, 1, 0)	(0, 0, 1),(1, 0, 0)	(1, 0, 0),(1, 0, 0)	(1, 0, 0),(0, 1, 0)	(0, 0, 0),(0, 1, 1)	(0, 0, 0),(1, 1, 0)
LGD	0.294	0.798	0.95	0.766	0.911	0.56
X_0	0.004	0.037	0.049	0.001		
Y_0	-0.031	-0.038	-0.056	0.021	[-0.046, -0.071]	[-0.023, -0.017]
a_m	[0]	[0]	[1]	[1]	[]	[]
α	0.577	1.335	2.601	0.146		
b_m	[1]	[1]	[1]	[1]	[]	[]
c_i	[1]	[1]	[1]	[1]	[1, 0]	[1, 1]
d_i	[0]	[1]	[1]	[0]	[0, 1]	[1, 0]
δ	0.017	0.024	0.004	-0.016	0.017	0.017
k	0.303	0.276	0.736	2.761	[0.596, 2.63]	[0.316, 2.403]
ν	0.001	0.032	0.01	0.17	[0.007, 0.151]	[0.016, 0.141]
ρ					-0.594	-0.896
σ	0.014	0.066	0.212	0.234		
spread	0.001	0.007	0.013	0.001	0.023	0.025
θ	0.022	0.035	0.012	0.06		

Table B.8: Germany and Italy

	(0, 0, 1),(0, 1, 0)	(0, 0, 1),(1, 0, 0)	(1, 0, 0),(1, 0, 0)	(1, 0, 0),(0, 1, 0)	(0, 0, 0),(0, 1, 1)	(0, 0, 0),(1, 1, 0)
LGD	0.354	0.583	0.665	0.605	0.37	0.328
X_0	0.003	0.083	0.001	0.001		
Y_0	-0.028	-0.042	-0.006	-0.018	[-0.028, -0.022]	[-0.022, -0.063]
a_m	[0]	[0]	[1]	[1]	[]	[]
α	0.87	1.809	0.04	0.591		
b_m	[1]	[1]	[1]	[1]	[]	[]
c_i	[1]	[1]	[1]	[1]	[1, 0]	[1, 1]
d_i	[0]	[1]	[1]	[0]	[0, 1]	[1, 0]
δ	0.014	0.013	-0.011	-0.002	0.014	0.023
k	0.237	0.485	1.88	0.175	[0.237, 0.883]	[1.803, 0.631]
ν	0.001	0.005	0.005	0.001	[0.001, 0.001]	[0.063, 0.087]
ρ					-0.348	0.012
σ	0.111	0.115	0.18	0.023		
spread	0.001	0.024	0.013	0.009	0.026	0.027
θ	0.025	0.002	0.094	0.016		

Table B.9: Germany and France

	(0, 0, 1),(0, 1, 0)	(0, 0, 1),(1, 0, 0)	(1, 0, 0),(1, 0, 0)	(0, 0, 0),(0, 1, 1)	(0, 0, 0),(1, 1, 0)
LGD	0.817	0.891	1	0.596	0.263
X_0	0.001	0.01	0.008		
Y_0	-0.039	-0.027	-0.016	[-0.038, -0.011]	[-0.008, -0.031]
a_m	[0]	[0]	[1]	[]	[]
α	0.116	0.001	0.029		
b_m	[1]	[1]	[1]	[]	[]
c_i	[1]	[1]	[1]	[1, 0]	[1, 1]
d_i	[0]	[1]	[1]	[0, 1]	[1, 0]
δ	0.029	0.019	-0.009	0.029	0.023
k	0.116	0.127	2.98	[0.103, 1.38]	[2.307, 0.105]
ν	0.018	0.008	0.25	[0.016, 0.034]	[0.094, 0.016]
ρ				-0.978	-0.853
σ	0.008	0.163	0.085		
spread	0	0.013	0	0.003	0.004
θ	0.007	0.022	0.079		

B.9. Dynamic Euribor calibration without credit risk

Figure B.26: Relative errors in Euribor fitting by maturities.

REFERENCES

Artzner, P. and Delbaen, F. (1995). "Default risk insurance and incomplete markets". *Mathematical Finance* 5(3), 187–195.

Bachelier, L. (1900). "Théorie de la spéculation". *Annales Scientifiques de l'École Normale Supérieure* 17(3), 21–86.

Bergman, Y. Z. (1982). "Pricing of contingent claims in perfect and in imperfect markets". PhD thesis. University of California, Berkeley.

Bielecki, T. R. and Rutkowski, M. (2002). *Credit Risk: Modeling, Valuation and Hedging*. 2nd. Berlin: Springer.

Billingsley, P. (1992). *Probability and Measure*. 3rd. Hoboken: John Wiley & Sons.

BIS (2015). Review of the Credit Valuation Adjustment (CVA) risk framework. URL: `https://www.bis.org/bcbs/publ/d325.pdf`. Cited 02.03.2019.

BIS (2018). BIS statistics. URL: `https://www.bis.org/statistics/index.htm`. Cited 27.04.2019.

Björk, T. (2004). *Arbitrage Theory in Continuous Time*. Oxford.

Black, F. (1976). "The pricing of commodity contracts". *Journal of financial economics* 3(1-2), 167–179.

Black, F. and Scholes, M. (1973). "The pricing of options and corporate liabilities". *The journal of political economy* 81(3), 637–654.

Brace, A., Gatarek, D., and Musiela, M. (1997). "The market model of interest rate dynamics". *Mathematical finance* 7(2), 127–155.

Brigo, D. and Mercurio, F. (2001). "A deterministic–shift extension of analytically–tractable and time–homogeneous short–rate models". *Finance and Stochastics* 5(3), 369–387.

Brigo, D. and Mercurio, F. (2007). *Interest rate models-theory and practice: with smile, inflation and credit*. 2nd. Berlin: Springer Science & Business Media.

Brown, R. H. and Schaefer, S. M. (1994). "Interest rate volatility and the shape of the term structure". *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 347(1684), 563–576.

Byrd, R. H. et al. (1995). "A limited memory algorithm for bound constrained optimization". *SIAM Journal on Scientific Computing* 16(5), 1190–1208.

Carr, P. and Madan, D. (1999). "Option valuation using the fast Fourier transform". *Journal of computational finance* 2(4), 61–73.

Chapman, D., Long, J., and Pearson, N. (1999). "Using Proxies for the Short Rate: When Are Three Months Like an Instant?" *The Review of Financial Studies* 12(4), 763–806.

Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton university press.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). "A theory of the term structure of interest rates". *Econometrica: Journal of the Econometric Society*, 385–407.

Croucher, M. (2014–). *Python versions of nearest correlation matrix algorithm.* URL: https://github.com/mikecroucher/nearest_correlation/. Cited 02.03.2019.

Delbaen, F. and Schachermayer, W. (1994). "A general version of the fundamental theorem of asset pricing". *Mathematische annalen* 300(1), 463–520.

Delbaen, F. and Schachermayer, W. (1998). "The fundamental theorem of asset pricing for unbounded stochastic processes". *Mathematische annalen* 312(2), 215–250.

Dothan, U. (1978). "On the term structure of interest rates". *Journal of Financial Economics* 6(1), 59–69.

Duffie, D. (2010). *Dynamic asset pricing theory*. Princeton University Press.

Duffie, D. and Kan, R. (1994). "Multi-factor term structure models". *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 347(1684), 577–586.

Duffie, D. and Kan, R. (1996). "A yield-factor model of interest rates". *Mathematical finance* 6(4), 379–406.

Eberhart, R. and Kennedy, J. (1995). "Particle swarm optimization". *Proceedings of the IEEE international conference on neural networks*. Vol. 4. Citeseer, pp. 1942–1948.

Fama, E. F. (1965). "The behavior of stock-market prices". *The journal of Business* 38(1), 34–105.

Filipović, D. (2009). *Term-Structure Models: A Graduate Course*. Berlin: Springer.

Geman, H., El Karoui, N., and Rochet, J.-C. (1995). "Changes of numeraire, changes of probability measure and option pricing". *Journal of applied Probability*, 443–458.

Hagan, P. S. et al. (2002). "Managing smile risk". *The Best of Wilmott* 1, 249–296.

Harrison, J. M. and Kreps, D. M. (1979). "Martingales and arbitrage in multiperiod securities markets". *Journal of Economic theory* 20(3), 381–408.

Harrison, J. M. and Pliska, S. R. (1981). "Martingales and stochastic integrals in the theory of continuous trading". *Stochastic processes and their applications* 11(3), 215–260.

Harrison, J. M. and Pliska, S. R. (1983). "A stochastic calculus model of continuous trading: complete markets". *Stochastic processes and their applications* 15(3), 313–316.

Heath, D., Jarrow, R., and Morton, A. (1990). "Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation". *Journal of Financial and Quantitative Analysis* 25(04), 419–440.

Heath, D., Jarrow, R., and Morton, A. (1992). "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation". *Econometrica* 60(1), 77–105.

Heston, S. L. (1993). "A closed-form solution for options with stochastic volatility with applications to bond and currency options". *Review of financial studies* 6(2), 327–343.

Higham, N. J. (2002). "Computing the nearest correlation matrix—a problem from finance". *IMA journal of Numerical Analysis* 22(3), 329–343.

Higham, N. J. (2013). *The Nearest Correlation Matrix*. URL: https://nickhigham.wordpress.com/2013/02/13/the-nearest-correlation-matrix/. Cited 30.03.2019.

Ho, T. S. and Lee, S.-B. (1986). "Term structure movements and pricing interest rate contingent claims". *the Journal of Finance* 41(5), 1011–1029.

Hull, J. C. and White, A. D. (1994). "Numerical procedures for implementing term structure models II: Two-factor models". *The Journal of Derivatives* 2(2), 37–48.

Hull, J. and White, A. (1990). "Pricing interest-rate-derivative securities". *The Review of Financial Studies* 3(4), 573–592.

International Settlements, B. for (2010). *BIS Quarterly Review, December 2010*. URL: `https://www.bis.org/publ/qtrpdf/r_qt1012.pdf`. Cited 03.11.2013.

International Swaps and Derivatives Association, Inc (2009). *2009 ISDA Credit derivates determinations committees and auction and settlement CDS protocol*. URL: `https://www.isda.org/a/WS6EE/Big-Bang-Protocol.pdf`. Cited 31.10.2013.

J. Schönbucher, P. (Jan. 2001). "A Libor Market Model with Default Risk". *SSRN Electronic Journal*.

James, J. and Webber, N. (2000). *Interest Rate Modelling*. Chichester: John Wiley & Sons.

Jamshidian, F. (1989). "An Exact Bond Option Formula". *The Journal of Finance* 44(1), 205–209.

Jamshidian, F. (1997). "LIBOR and swap market models and measures". *Finance and Stochastics* 1(4), 293–330.

Jarrow, R. A. and Turnbull, S. M. (1995). "Pricing derivatives on financial securities subject to credit risk". *The journal of finance* 50(1), 53–85.

Jones, E., Oliphant, T., Peterson, P., et al. (2001–). *SciPy: Open source scientific tools for Python*. URL: `http://www.scipy.org/`.

Jovanovic, F. and Le Gall, P. (2001). "Does God practice a random walk? The 'financial physics' of a nineteenth-century forerunner, Jules Regnault". *European Journal of the History of Economic Thought* 8(3), 332–362.

Lando, D. (1998). "On Cox processes and credit risky securities". *Review of Derivatives research* 2(2-3), 99–120.

Litterman, R. and Scheinkman, J. (1991). "Common factors affecting bond returns". *Journal of fixed income* 1(1), 54–61.

McNeil, A. J., Frey, R., and Embrechts, P. (2010). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton: Princeton university press.

Mercurio, F. (2009). "Interest rates and the credit crunch: new formulas and market models". *Bloomberg portfolio research paper* (2010-01).

Merton, R. C. (1971). "Optimum consumption and portfolio rules in a continuous-time model". *Journal of economic theory* 3(4), 373–413.

Merton, R. C. (1974). "On the pricing of corporate debt: The risk structure of interest rates". *The Journal of Finance* 29(2), 449–470.

Miltersen, K. R., Sandmann, K., and Sondermann, D. (1997). "Closed form solutions for term structure derivatives with log-normal interest rates". *The Journal of Finance* 52(1), 409–430.

Musiela, M. and Rutkowski, M. (2005). *Martingale Methods in Financial Modelling*. Milan: Springer.

Nawalka, S., Beliaeva, N., and Soto, G. (2007). *Dynamic Term Structure Modeling: The Fixed Income Valuation Course*. Hoboken, New Jersey: John Wiley & Sons.

O'Kane, D. (2011). *Modelling Single-name and Multi-name Credit Derivatives*. Chichester: John Wiley & Sons.

Øksendal, B. (2003). *Stochastic differential equations*. Berlin: Springer.

Pye, G. (1966). “A Markov model of the term structure”. *The Quarterly Journal of Economics*, 60–72.

Regnault, J. (1863). *Calcul des chances et philosophie de la Bourse*.

Ritchken, P. and Sankarasubramanian, L. (1995). “Volatility structures of forward rates and the dynamics of the term structure”. *Mathematical finance* 5(1), 55–72.

Rytty, M. (2019). *GitHub-repository for the code*. URL: `https://github.com/mrytty/gradu-public/`.

Samuelson, P. A. (1965a). “Proof that properly anticipated prices fluctuate randomly”. *Industrial management review* 6(2).

Samuelson, P. A. (1965b). “Rational theory of warrant pricing”. *Industrial management review* 6, 13–31.

Samuelson, P. A. (1973). “Mathematics of speculative price”. *SIAM Review* 15(1), 1–42.

Storn, R. (1996). “On the usage of differential evolution for function optimization”. *Proceedings of North American Fuzzy Information Processing*. IEEE, pp. 519–523.

Storn, R. and Price, K. (1997). “Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces”. *Journal of global optimization* 11(4), 341–359.

Vašíček, O. (1977). “An equilibrium characterization of the term structure”. *Journal of financial economics* 5(2), 177–188.

Wu, L. (2009). *Interest rate modeling: Theory and practice*. Chapman and Hall/CRC.

Zhu, C. et al. (1997). “Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization”. *ACM Transactions on Mathematical Software (TOMS)* 23(4), 550–560.